

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 15:47:13 ; Search time 15.22 Seconds
(without alignments)
2599.959 Million cell updates/sec

Title: US-09-049-304A-122
Perfect score: 5286
Sequence: 1 CARLLGGGKNGPRVNRIV.....VPALEILSSGKLVKVEVET 1022

Scoring table:

BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	672.5	12.7	446	1	LYS9_YEAST
2	172	3.3	369	1	LYS1_YARLI
3	157	3.0	368	1	LYS1_SCHPO
4	153.5	2.9	373	1	LYS1_YEAST
5	144.5	2.7	382	1	LYS1_CANAL
6	141.5	2.7	1861	1	MAP2_RAT
7	139.5	2.6	604	1	DNAB_BACME
8	125.5	2.4	1451	1	SPT6_YEAST
9	124	2.3	5255	1	BACA_BACLI
10	121	2.3	2672	1	GCNL_YEAST
11	120	2.3	493	1	CETP_HUMAN
12	119	2.3	1828	1	MAP2_MOUSE
13	118	2.2	428	1	ENO_2YMMO
14	116.5	2.2	1450	1	RPO1_ASFB7
15	116.5	2.2	4644	1	DYHC_MOUSE
16	115.5	2.2	493	1	AMVR_DROTE
17	115.5	2.2	494	1	AMVR_DROBP
18	115.5	2.2	562	1	EZPA_BACSU
19	115.5	2.2	680	1	LAG2_YEAST
20	115.5	2.2	3587	1	TYCB_BACBR
21	115	2.2	858	1	YNU1_YEAST
22	114.5	2.2	3176	1	CA36_HUMAN
23	114	2.2	890	1	POL2_BAYMG
24	114	2.2	1416	1	RPOC_TREPA
25	113.5	2.1	986	1	GUNZ_CLOSAR
26	113	2.1	1244	1	DROL_ASFL6
27	112.5	2.1	313	1	DICO_LACLA
28	112.5	2.1	626	1	HCY6_ANDAU
29	112.5	2.1	2863	1	LRBA_HUMAN
30	112	2.1	1862	1	ANK1_MOUSE
31	111.5	2.1	493	1	AMVR_DROSE
32	111.5	2.1	1023	1	RT12_ACTPL
33	111.5	2.1	1679	1	YIO9_YEAST

34	111	2.1	1240	1	DP3A_CHLPN
35	110	2.1	1132	1	PHY1_PHYPA
36	109.5	2.1	493	1	AMVR_DROER
37	109.5	2.1	493	1	AMVR_DROME
38	109.5	2.1	610	1	DNAB_BACSH
39	109.5	2.1	1595	1	SOS_DROME
40	109	2.1	1024	1	POPC_RALSO
41	109	2.1	1592	1	YNY2_YEAST
42	108.5	2.1	493	1	AMVR_DROYA
43	108.5	2.1	511	1	ATPB_YEAST
44	108.5	2.1	977	1	A2A1_MOUSE
45	108.5	2.1	984	1	SECA_AQUAE

Q927n8 chlamydia p
P36505 physcomitre
O76265 drosophila
O18408 drosophila
O69268 bacillus sp
P26675 drosophila
Q9rbu2 ralstonia s
P53855 saccharomyc
O76264 drosophila
P00830 saccharomyc
P17426 mus musculu
O67718 aquifex aeo

ALIGNMENTS

RESULT 1
LYS9_YEAST
ID LYS9_YEAST STANDARD; PRT; 446 AA.
AC P38999;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Saccharopine dehydrogenase [NADP+, L-glutamate forming] (EC 1.5.1.10).
GN LYS9 OR LYS13 OR YNR050C OR N3461.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=1278B;
RA Feller A.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RA Pohl T.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE OF 314-324.
RC STRAIN=ATCC 38531 / Y41;
RX MEDLINE=97089742; PubMed=8935650;
RA Norbeck J., Blomberg A.;
RT "Protein expression during exponential growth in 0.7 M NaCl medium of
Saccharomyces cerevisiae.";
RL FEMS Microbiol. Lett. 137:1-8(1996).
CC -1- CATALYTIC ACTIVITY: N6-(L-1,3-dicarboxypropyl)-L-lysine + NADP(+) + H(2)O = L-glutamate + 2-aminoadipate-6-semialdehyde + NADPH.
CC -1- PATHWAY: SEVENTH STEP IN LYSINE BIOSYNTHESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X77363; CAAS4552.1; -
DR EMBL; 271665; CAAG6331.1; -
DR PIR; S41937; S41937.
DR YEPD; 8416; -
DR SGD; S0005333; LYS9.
KW Lysine biosynthesis; Oxidoreductase; NADP.
SQ SEQUENCE 446 AA; 48917 MW; 1EC4CEICE3BCD916 CRC64;

Query Match 12.7%; Score 672.5; DB 1; Length 446;
Best Local Similarity 35.1%; Pred. No. 3.5e-36;
Matches 170; Conservative 82; Mismatches 183; Indels 49; Gaps 13;
QY 543 GPKILILGRCVCPAAEFLLASYPDICTYGVDDHDADQIHVIVASLYQKDAETVDGIEN 602

Db	2	GNVLLGSGFVAQPV	IDLAA-----NDINVTVACHTIANQAALAKP-SG	47
QY	603	TTATOLDVADIGSLDSLVQSVQVVISILLPASFAHAAIAGVCIELKHKHWITASTVDESMSNL	662	
Db	48	SKAISLDVTDGSDALDKVLADNDVVISLIPYTFHPNVVKSARTKTDVVTSSYISPALREL	107	
QY	663	SQAAKDAGVTIILCEMLDGPIDHLMWSKMDIEAHARKGKTKAFTSYCGGLPSPAAANPL	722	
Db	108	EPEIVKAGITVMEIGLDPGIDHLYAVKTIDEVHARAGGKLUKSLFSCGGLPAPEDSDNPL	167	
QY	723	AYKFSNPAGALRSKGNPAVYKFLGETTHVDCHNLYESAKRLRLRELPALEHLNPNRS	782	
Db	168	GYKFSWSSRGVLLALNRNSKYWKDGKIETVSSDLMATAKPYFI--YPGYAFVCIYENRDS	225	
QY	783	LIYGDLYGISKEASTIYRATXRYEGFSEIMVTLKGTFF-DAGNH-----PLLQDTSRPTV	837	
Db	226	TLFKDLYHI-PEAETVIRGTLRYOGFPEFVKALVDMGLKDDANEIYSFKPIANNEALKQY	284	
QY	838	KGFUDELNNISTINTDLDIEASGGYDDDLIARLLKLGCCRNKEIAVYKVTIKFGLGHE	897	
Db	285	LG-----AKSYSKEDLIASIDSKATWKDDREIRILSGFAWLGLFS	325	
QY	898	ETQI-PRKGCSSPFDVTCORMEORMAYGHNEQDMVLLHHEVEVEYDQOPAEKHQATLLEF	956	
Db	326	DAKITPRG--NALDITLCARLEEMOYEDNEROMVQLQKFGIEWADG-TTETRTSTLVDY	382	
QY	957	GVKENGSRSTAMATVTCIPAAIGALLLNKKVQTKGVRIPQIPEIYVVPAL-EILESSGK	1015	
Db	383	KKV--GGYS--SMAATVGVPAIATKFKVLDGTIKGPGLLAPYSPEINDPMKELKDKYGIY	439	
QY	1016	LVEK 1019		
Db	440	LXEK 443		
RESULT	2			
LYSL_YARLI	ID	LYSL_YARLI	STANDARD; PRT; 369 AA.	
AC	DT	01-FEB-1995 (Rel. 31, Created)		
DT	DT	01-FEB-1995 (Rel. 31, Last annotation update)		
DT	DT	15-DEC-1998 (Rel. 37, Last annotation update)		
DE	DE	Saccharopine dehydrogenase [NAD+, L-lysine forming] (EC 1.5.1.7)		
DE	DE	(lysine--2-oxoglutarate reductase) (SDH).		
GN	OS	Yarrowia lipolytica (Candida lipolytica).		
OC	OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OX	OX	Saccharomycetales; Dipodascaceae; Yarrowia.		
RN	RN	NCBI_TaxID=4952;		
RP	RP	SEQUENCE FROM N.A.		
RC	RC	STRAIN=ATCC 20460 / W29;		
RX	RX	MEDLINE=90355936; PubMed=2386625;		
RA	RA	Xuan J.-W., Fournier P., Declercq N., Charles M., Gaillardin C.;		
RT	RT	"Overlapping reading frames at the LYS5 locus in the yeast Yarrowia		
RT	RT	lipolytica";		
RL	RL	Mol. Cell. Biol. 10:4795-4806(1990).		
CC	CC	-1- CATALYTIC ACTIVITY: N6-(L-1,3-dicarboxypropyl)-L-lysine + NAD(+) +		
CC	CC	H2O) = L-lysine + 2-oxoglutarate + NADH.		
CC	CC	-1- PATHWAY: 8TH STEP IN LYSINE BIOSYNTHESIS.		
CC	CC	-1- SIMILARITY: TO OTHER FUNGAL SACCAROPINE DEHYDROGENASE.		
CC	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	CC	use by non-profit institutions as long as its content is in no way		
CC	CC	modified and this statement is not removed. Usage by and for commercial		
CC	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	CC	or send an email to license@isb-sib.ch).		
DR	DR	EMBL; M34929; AAA35248.1;		
DR	DR	PIR; A36467; A36467.		

KW	Lysine biosynthesis; Oxidoreductase; NAD.
FT	ACT_SITE 205 BY SIMILARITY.
SQ	SEQUENCE 369 AA; 40612 MW; B1B749FA008BA36 CRC64;
	Query Match 3.3%; Score 172; DB 1; Length 369;
	Best Local Similarity 20.5%; Pred. No. 0.00063;
Matches	81; Conservative 64; Mismatches 159; Indels 92; Gaps 12;
QY	18 IIVQPSTRIHDAOYEDAGCEISDLS-----ECGLIIGIKQPKLOMILSDRAYAFESH 72
DB	37 VFVEKSLRIFDDQFVDVGTALVEEGSWYSAPEDRMIIGLKELPESPFLSHEHIQHAFH 96
QY	73 THRAAKENMLDKTILERSVLFDYELIVGDGKRSLAFGKFAGRAGLIDFLHLGQRYL 132
DB	97 CYKDQGWDKLVLSRFPAGNGTLIDLEDDNGRRVAFGFHAGFAG-----A 144
QY	133 SLGVSTPLSLGQSH-----MYPSLAANKAAVVVAEIAINFGLPSGICPIVFVF 182
DB	145 AIGVET--WAFQOTHDPDENLPGSVAYPN---ETELVDKKTKDLAAAVEKGSKLPTVLVI 199
QY	183 TGVGNVSQGAQEITFKLLPTHFTVDAEKLPFIQAANLSKQSOSTKRVFYLGCVVTSRDIV 242
DB	200 GALGRCSGGAIDLARKV-----GIPE---ENLRWDMNETK-----232
QY	243 SHKPOTRFQDGKYAYAHPEHYTFVHERIATPYASVINVCYWKEKREPPMLNMQOLQMLE 302
DB	233 -----KGGPF-----QETAD-ADIFNICITLSQIPPIFFNYDLLNKETR 270
QY	303 TCCLPLYGCDITCDIGGSIEFIINKSTSIERRPFYDPDSKNSYHDMDEAGVGVCVLDILP 362
DB	271 KLSIVDVSDATTNPHNPVPVYTIIATTDHPATVPVETT-----AGPKLSVCSIDHLP 322
QY	363 TEFSKEASQHGNTLSRLVASLASVKQPAELPSYLR 398
DB	323 SLLPREASEAFS----EALLPSLLQLQFOROTAPVWTR 355
Nb	

```

RESULT 3
ID LYS1_SCHPO STANDARD; PRF; 368 AA.
ID LYS1_SCHPO
AC Q09694; Q9UTC1;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Seccharopine dehydrogenase [NAD+, L-lysine forming] [EC 1.5.1.7]
DE (Lysine - 2-oxoglutarate reductase) (SDH).
DE (Lysine - 2-oxoglutarate reductase) (SDH).
DE LYS3 OR SPAC227.18 OR SPAC2F7 01.
DE Schizosaccharomyces pombe (Fission yeast).
OS Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID:4896;
[1]
RN SEQUENCE FROM N.A.
RA STRAIN=972;
RC Zimmermann W., Wambutt R., McDougall R.C., Rajandream M.A.,
RA Barrell B.G.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE OF 320-368 FROM N.A.
RC STRAIN=972;
RA Gentles S., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -! CATALYTIC ACTIVITY: N6-(L-1,3-dicarboxypropyl)-L-lysine + NAD(+) +
CC H(2)O -> L-lysine + 2-oxoglutarate + NADH.
CC -! PATHWAY: 8TH STEP IN LYSINE BIOSYNTHESIS.
CC -! SIMILARITY: TO OTHER FUNGAL SACCHAROPINE DEHYDROGENASE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no

```


Tue Jun 4 15:23:16 2002

```

GN  MAP2.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.
RC  STRAIN=WISTAR; TISSUE=Brain;
RX  MEDLINE=90251471; PubMed=2339070;
RA  Kindler S., Schwanke B., Schulz B., Garner C.C.;
RT  "Complete cDNA sequence encoding rat high and low molecular weight
    MAP2.";
RL  Nucleic Acids Res. 18:2822-2822(1990).
RN  [2]
RP  SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.
RC  STRAIN=WISTAR; TISSUE=Brain;
RX  MEDLINE=90251471; PubMed=2339070;
RA  Kindler S., Schwanke B., Goedert M., Garner C.C.;
RT  "Molecular structure of microtubule-associated protein 2b and 2c from
    rat brain.";
RL  J. Biol. Chem. 265:19679-19684(1990).
RN  [3]
RP  SEQUENCE OF 1-1511; 1515-1694 AND 1726-1861 FROM N.A.
RC  MEDLINE=90221819; PubMed=2326166;
RA  Doll T., Papadrikopoulou A., Matus A.;
RT  "Nucleotide and amino acid sequences of embryonic rat MAP2c.";
RL  Nucleic Acids Res. 18:361-361(1990).
RN  [4]
RP  DISCUSSION OF SEQUENCE.
RC  MEDLINE=89365159; PubMed=2770869;
RA  Papadrikopoulou A., Doll T., Tucker R.P., Garner C.C., Matus A.;
RT  "Embryonic MAP2 lacks the cross-linking sidearm sequences and
    dendritic targeting signal of adult MAP2.";
RL  Nature 340:650-652(1989).
RN  [5]
RP  SEQUENCE OF 1695-1725 FROM N.A.
RC  MEDLINE=94110302; PubMed=8282767;
RA  Doll T., Meichsner M., Riederer B.M., Honegger P., Matus A.;
RT  "An isoform of microtubule-associated protein 2 (MAP2) containing
    four repeats of the tubulin-binding motif.";
RL  J. Cell Sci. 106:633-640(1993).
CC  -!- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY
    STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO
    SEEM TO HAVE A STIFFENING EFFECT ON MICROTUBULES.
CC  -!- ALTERNATIVE PRODUCTS: VARIOUS FORMS OF MAP2 ARE PRODUCED BY
    ALTERNATIVE SPLICING OF THE SAME GENE. MAP2C, THE LOW MOLECULAR
    FORM OF MAP2, LACKS THE CENTRAL DOMAIN OF MAP2A/B.
CC  -!- DEVELOPMENTAL STAGE: MAP2C IS EXPRESSED DURING EMBRYONIC BRAIN
    DEVELOPMENT AND UNTIL POSTNATAL DAY 10. MAP2B IS EXPRESSED
    THROUGHOUT BRAIN DEVELOPMENT.
CC  -!- SIMILARITY: CONTAINS 3 OR 4 TAU/MAP REPEATS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL: X51842; CAA36135.1; -
CC  EMBL: X17682; CAA35667.1; -
CC  EMBL: X71487; CAA50588.1; -
CC  PIR: S07887; S07887.
CC  PIR: S10003; S10003.
CC  PIR: A37981; A37981.
CC  InterPro: IPR001084; Tubulin-bind.
CC  Pfam: PF00418; tubulin-binding; 4.
CC  PROSITE: PS00229; TAU_MAP; 3.
CC  Microtubules; Repeat; Alternative splicing; Calmodulin-binding.
CC  DOMAIN 1454 1474 CALMODULIN-BINDING (POTENTIAL).
CC  REPEAT 1664 1694 TAU/MAP MOTIF.
CC  REPEAT 1695 1725 TAU/MAP MOTIF.
CC  -----

```

```

GN  LYS1.
OS  Candida albicans (Yeast).
OC  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC  Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX  NCBI_TaxID=5476;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  MEDLINE=95012722; PubMed=7927784;
RA  Garrad R.C., Schmidt T.M., Bhattacharjee J.K.;
RT  "Molecular and functional analysis of the LYS1 gene of Candida
    albicans.";
RL  Infect. Immun. 62:5027-5031(1994).
RN  [2]
RP  CATALYTIC ACTIVITY: N6-(L-L-3-dicarboxypropyl)-L-lysine + NAD(+) +
    H2O = L-lysine + 2-oxoglutarate + NADH.
CC  -!- PATHWAY: 8TH STEP IN LYSINE BIOSYNTHESIS.
CC  -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC  -!- SIMILARITY: TO OTHER FUNGAL SACCAROPINE DEHYDROGENASE.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL: U13233; AAA21362.1; -
CC  Lysine biosynthesis; Oxidoreductase; NAD.
CC  ACT_SITE 217 217 BY SIMILARITY.
CC  SEQUENCE 382 AA; 42394 MW; A3620191DF04B88F CRC64;
CC  -----
Query Match 2.7%; Score 144.5; DB 1; Length 382;
Best Local Similarity 21.2%; Pred. No. 0.04;
Matches 88; Conservative 68; Mismatches 161; Indels 99; Gaps 18;
QY 18 IVQPSSTRIHDAQYEDAGCEISDELSDLS-----ECGLIIGIKQ-PKQLMLSDRAVAFPS 71
DB 18 IVQPSSTRIHDAQYEDAGCEISDELSDLS-----ECGLIIGIKQ-PKQLMLSDRAVAFPS 71
QY 38 IYVESSQSTFDIKYEAVGAKIYVEGSKWTKPRKRIIFGLKPELTETPLIHEIHQIA 97
DB 38 IYVESSQSTFDIKYEAVGAKIYVEGSKWTKPRKRIIFGLKPELTETPLIHEIHQIA 97
QY 72 HTYHAKQENKMLKILREVSFLFYELIVGDGKRSIAFGKAFRA-----GLIDF---- 123
DB 72 HTYHAKQENKMLKILREVSFLFYELIVGDGKRSIAFGKAFRA-----GLIDF---- 123
QY 98 HCYKDAQWQDVLKRPQNGILYDLENDQGRRAAFYAGFYAGAGAGLVLDWFSFKQ 157
DB 98 HCYKDAQWQDVLKRPQNGILYDLENDQGRRAAFYAGFYAGAGAGLVLDWFSFKQ 157
QY 124 LHG--LGQYLSIGYSTPLSLGQSHMYPSLAAKAAVIVVAEATFGLPSGICPIVFV 181
DB 124 LHG--LGQYLSIGYSTPLSLGQSHMYPSLAAKAAVIVVAEATFGLPSGICPIVFV 181
QY 158 LNGTKTGKGEGGGLPGVT-----PYPNELIKDKVIELEKALTKN--GGQYPRCLV 210
DB 158 LNGTKTGKGEGGGLPGVT-----PYPNELIKDKVIELEKALTKN--GGQYPRCLV 210
QY 182 FTGVGNVSGQAQEIFKLLPHTFVDAEKLPEITFOARNLSKQSTKRVFLYGCVTSRDI 241
DB 182 FTGVGNVSGQAQEIFKLLPHTFVDAEKLPEITFOARNLSKQSTKRVFLYGCVTSRDI 241
QY 211 IGAIGRCGSGAIDLEFKKI-----GIPD-----DNIARWMA----- 241
DB 211 IGAIGRCGSGAIDLEFKKI-----GIPD-----DNIARWMA----- 241
QY 242 VSHKDPTRQFDKGYAHPEHYTPVPHERAPYASVIVNMYWKRFPPLNMOLOQLM 301
DB 242 VSHKDPTRQFDKGYAHPEHYTPVPHERAPYASVIVNMYWKRFPPLNMOLOQLM 301
QY 242 -----ETAKG-----PQEIYD--SDIFINCYLSKPIPPINKELTN--- 278
DB 242 -----ETAKG-----PQEIYD--SDIFINCYLSKPIPPINKELTN--- 278
QY 302 ETGCPYGVCDITCDIGGSIEFINKSTSIERPFRY-----DPSKNSYHDDMEGAVV 354
DB 302 ETGCPYGVCDITCDIGGSIEFINKSTSIERPFRY-----DPSKNSYHDDMEGAVV 354
QY 279 NENRKLITIVDSAD-----TTNPNHPIPVYEIATVNEPTVEVKLD--KGPKL 326
DB 279 NENRKLITIVDSAD-----TTNPNHPIPVYEIATVNEPTVEVKLD--KGPKL 326
QY 355 CLAVDILPTESKASQHFNGNLSRLVASLKVQPAELPSYLRACI--AHAGRL 408
DB 355 CLAVDILPTESKASQHFNGNLSRLVASLKVQPAELPSYLRACI--AHAGRL 408
QY 327 VCSIDHLPSPREASEFFA---KDLMPSLLELPNDRTPSVVVRKAKQLEFKHVARL 379
DB 327 VCSIDHLPSPREASEFFA---KDLMPSLLELPNDRTPSVVVRKAKQLEFKHVARL 379
RESULT 6
MAP2_RAT
ID MAP2_RAT
AC P15146;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 2 (MAP 2) (MAP2B) [Contains: MAP2C].

```


QY 961 NGRSTTAMALTVGIPAAIGALLKKNKQVTKVIRPLOPEIYVPALEILESSGIKLVKV 1020
 Db 1057 -----IELLRDDPRRAKLESNLSEYAELE--KNTGLRKLNNL 1094
 QY 1021 ET 1022
 Db 1095 NT 1096

RESULT 9
 ID BACA_BACLI
 AC O68006; STANDARD; PRT; 5255 AA.
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bacitracin synthetase 1 (BAC1) [Includes: ATP-dependent isoleucine
 DE adenylation (ileA) (Isoleucine activase); ATP-dependent cysteine
 DE adenylation (CysA) (Cysteine activase); ATP-dependent leucine adenylation
 DE (LeuA) (Leucine activase); ATP-dependent glutamate adenylation (GluA)
 DE (Glutamate activase); ATP-dependent isoleucine adenylation (IleA)
 DE (Isoleucine activase); Glutamate racemase (EC 5.1.1.3)].
 GN BACA.
 OS Bacillus licheniformis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID-1402;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 10716;
 RX MEDLINE-98089193; PubMed-9427658;
 RA Konz D., Klenz A., Schoergendorfer K., Marahiel M.A.;
 RT "The bacitracin biosynthesis operon of Bacillus licheniformis ATCC
 RT 10716: molecular characterization of three multi-modular peptide
 RT synthetases.";
 RL Chem. Biol. 4:927-937(1997).
 CC -!- FUNCTION: ACTIVATES FIVE AMINO ACIDS. INCORPORATES TWO D-AMINO
 CC ACIDS, RELEASES AND CYCLIZES THE MATURE BACITRACIN.
 CC -!- CATALYTIC ACTIVITY: L-glutamate - D-glutamate.
 CC -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
 CC (POTENTIAL).
 CC -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
 CC ANTIBIOTIC BACITRACIN.
 CC -!- SUBUNIT: LARGE MULTIZYME COMPLEX OF BAC1, BAC2 AND BAC3.
 CC -!- DOMAIN: CONSISTS OF FIVE MODULES AND ONE EPIMERIZATION DOMAIN IN
 CC THE FOURTH MODULE. EACH MODULE INCORPORATES ONE AMINO ACID INTO
 CC THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
 CC RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOLATION, CONDENSATION
 CC (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND
 CC N METHYLATION (OPTIONAL).
 CC -!- MISCELLANEOUS: BACITRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
 CC DODECAPEPTIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
 CC ABUNDANT IS BACITRACIN A, A BRANCHED CYCLIC DODECAPEPTIDE. IT
 CC CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-
 CC GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION
 CC PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-
 CC HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
 CC TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
 CC CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
 CC PHE-9, AND ASP-11).
 CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -!- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF007865; AAC06346.1; -

DR HSSP: P14687; 1AMU.
 DR InterPro: IPR000873; AMP-bind.
 DR InterPro: IPR001242; DUF4.
 DR InterPro: IPR003880; Phosphopant_attach.
 DR Pfam: PF00501; AMP-binding; 5.
 DR Pfam: PF00668; Condensation; 5.
 DR Pfam: PF00550; pp-binding; 5.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 3.
 DR PROSITE: PS00455; AMP_BINDING; 5.
 DR PROSITE: PS0075; ACP_DOMAIN; 5.
 KW Ligase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
 KW Multifunctional enzyme; Repeat
 FT REPEAT 39 612 DOMAIN 1 (ISOLEUCINE-ACTIVATING).
 FT REPEAT 1109 1648 DOMAIN 2 (CYSTEINE-ACTIVATING).
 FT REPEAT 2124 2889 DOMAIN 3 (LEUCINE-ACTIVATING).
 FT REPEAT 3164 3732 DOMAIN 4 (GLUTAMINE-ACTIVATING).
 FT REPEAT 4668 5249 DOMAIN 5 (ISOLEUCINE-ACTIVATING).
 FT DOMAIN 621 1037 CYCLIZATION (POTENTIAL).
 FT DOMAIN 544 611 ACYL CARRIER (ACP) 1.
 FT DOMAIN 1585 1652 ACYL CARRIER (ACP) 2.
 FT DOMAIN 2621 2688 ACYL CARRIER (ACP) 3.
 FT DOMAIN 3664 3730 ACYL CARRIER (ACP) 4.
 FT DOMAIN 5171 5238 ACYL CARRIER (ACP) 5.
 FT BINDING 574 574 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 1615 1615 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 2651 2651 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 3694 3694 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 5201 5201 PHOSPHOPANTETHEINE (BY SIMILARITY).
 SQ SEQUENCE 5255 AA; 598254 MW; 906E8DD68450F85B CRC64;

Query Match 2.3%; Score 124; DB 1; Length 5255;
 Best Local Similarity 18.5%; Pred. No. 46;
 Matches 141; Conservative 119; Mismatches 302; Indels 202; Gaps 37;

QY 44 LSEGLIGIKQPKLWILSDRAYAFSHT-----HKAQENMPLLD-KILERVSLF 95
 Db 2718 LSTGSHVAVNMPFAMTLEGDFDIRFENTLKNMVRHESFRTSFVMIDGVNQIENEI 2777
 QY 96 DYELIVDDGKRSALAFCKFAGRLDILHLGLQRLSLGYSYTPFLSCQSHYPS-LAA 154
 Db 2778 DFQVAYSDIGKES-----AEKIKSFTR-----HLEKAPLLR 2811
 QY 155 AKAAVIVVAEEIATFG----LPSGICPIVFTGVGNVSG-AQEIFKLLPHFTVDAEKL 209
 Db 2812 AEVVKLNREHLLMDFMHHIISDGVSTDFI-QELGALYEGKSLKPFHIQYKDYAEWE-- 2868
 QY 210 PEIFQARNLSKQSTKRVFQLYGCVVTSRDI-----VSHKDPTRQFDKGDYIAHPEHY 263
 Db 2869 -----NSHARSEELKQ-EEYWLKTYKGDIPLVDLPIDHKRPLTKSSEG- 2912
 QY 264 TPVHERIAPYASVIVNCMWKRPPLNMDQLOQLMTCPLVGVGCDITCDIGGSTEF 323
 Db 2913 -----TVTAAIESETERKLOHMAK-----ENGVWYMLL-----LAGYTAL 2948
 QY 324 INKSTSIERPFRYPDKSNVSHDDMEGAGVCLAVDILPTFEFSKEASOHFNILSRVAS 383
 Db 2949 LSKYTG-QEDIIVGTGPAARNHEDIQ--HLIGMEFVNTLAIRNHPKTRFDYLOEYKEN 3005
 QY 384 LASVQKPAELP--SYLRRACIAHAGRLTPLYEYIPRNRNTMIDLAPAKTNPLPKYSTL 441
 Db 3006 TLQAYENODYFEELVEKVNKRMARNPLFTMLVYHNT--DVKPEARGL-----RSRL 3059
 QY 442 VLSGLHFDKFLINEALDIITAGSFHLVRCVCGQSDTDDMSYSELEVAGDATTATLDKII 501
 Db 3060 VEIRKGI-SKF-----DITVTASEAADGURLEVEYSTTLFNKERME-----RLS 3102
 QY 502 DLSLSLANEHGGDHDAQOEIETALKIGKVNEYETIDKGGPKLILILGAGRCVCPAAEF 561
 Db 3103 EHLISLL-EQAADH-----PDATN-----QIDVLTKGERHVLDF 3138
 QY 562 LASYPDICTYGVDDHDADQIHVIVASLYKQDAETVD-----GIENTTATQLDVADIGS 615

GCN1_YEAST STANDARD; PRT: 2672 AA.

AC P33892; 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Translational activator GCN1.

DE GCN1 OR YGL195W OR G1318.

OS Saccharomyces cerevisiae (Baker's yeast).

OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93268304; PubMed=8497269;

RA Marton M.J., Crouch D., Hinnebusch A.G.;

RT "GCN1, a translational activator of GCN4 in Saccharomyces cerevisiae,

RT is required for phosphorylation of eukaryotic translation initiation

RT factor 2 by protein kinase GCN2.";

RT Mol. Cell. Biol. 13:3541-3556(1993).

RL [2]

RN SEQUENCE FROM N.A.

RC STRAIN=S288C / FY1679;

RX MEDLINE=97197971; PubMed=9046087;

RA Coglievina M., Klima R., Bertani I., Delneri D., Zaccaria P.,

RA Bruchi C.V.;

RT "Sequencing of a 40.5 kb fragment located on the left arm of

RT chromosome VII from Saccharomyces cerevisiae.";

RL Yeast 13:55-64(1997).

CC -1- FUNCTION: TRANSLATIONAL ACTIVATOR OF GCN4. MAY BE INVOLVED IN

CC SENSING CHARGED tRNA AND STIMULATING THE KINASE ACTIVITY OF GCN2

CC IN AMINO ACID-STARVED CELLS. REQUIRED IN VIVO FOR THE

CC PHOSPHORYLATION OF EIF-2-ALPHA ON SERINE-52 BY THE PROTEIN KINASE

CC GCN2.

CC -1- SUBUNIT: COMPONENT OF A HETEROMERIC COMPLEX THAT INCLUDES GCN1 AND

CC GCN20.

CC -1- SIMILARITY: STRONG, TO S.POMBE SPAC18G6.05C.

CC -1- SIMILARITY: CONTAINS 20 HEAT REPEATS.

CC

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; L12467; AAA34635.1; -

CC EMBL; X91837; CAA62949.1; -

CC EMBL; 272717; CAA96907.1; -

CC PIR; A48126; A48126.

CC SGD; S0003163; GCN1.

CC InterPro; IPR000357; HEAT_repeat.

CC PROSITE; PS00077; HEAT_REPEAT; 4.

CC Translation regulation; Activator; Repeat.

CC REPEAT 932 970 HEAT 1.

FT REPEAT 1030 1067 HEAT 2.

FT REPEAT 1099 1138 HEAT 3.

FT REPEAT 1243 1281 HEAT 4.

FT REPEAT 1284 1321 HEAT 5.

FT REPEAT 1405 1442 HEAT 6.

FT REPEAT 1484 1521 HEAT 7.

FT REPEAT 1523 1559 HEAT 8.

FT REPEAT 1561 1598 HEAT 9.

FT REPEAT 1603 1640 HEAT 10.

FT REPEAT 1641 1679 HEAT 11.

FT REPEAT 1721 1758 HEAT 12.

FT REPEAT 1760 1796 HEAT 13.

FT REPEAT 1862 1903 HEAT 14.

FT REPEAT 1905 1942 HEAT 15.

FT REPEAT 1947 1984 HEAT 16.

FT REPEAT 1985 2024 HEAT 17.

FT REPEAT 2097 2134 HEAT 18.

FT REPEAT 2290 2328 HEAT 19.

FT REPEAT 2347 2384 HEAT 20.

SQ SEQUENCE 2672 AA; 296693 MW; 980FDD03753E9D1C CRC64;

Query Match. 2.3%; Score 121; DB 1; Length 2672;

Best Local Similarity 18.4%; Pred. No. 25;

Matches 201; Conservative 156; Mismatches 399; Indels 334; Gaps 48;

QY 11 NGRVNRIVVOPSTRRIH-HDAQYEDAGCEISEDLSCGLIIGIKQ--PKLQMLSD--- 64

DB 1559 DGPSL-ALIIHIIHRGMHDSRANKRACKI--VGNMILVDTKLLIPVLOQLIDEVEI 1614

QY 65 -----RAYAFSFHTHKAKENMP-----LDDKILEERVSFLFDYELIVGDDGKR 107

DB 1615 AMVDVPVNTATARAALGALVERLGEQFPLIPRLDTLSDSKS-----GD----- 1662

QY 108 SLAFCKFAGRAGLIDFLHGLQRYLSIGYSTPFLSLGSHMYPSPSLAAKAAVIVAEIEA 167

DB 1663 -----RLGSAQALAEVIGSLGLTKL-----DEMLPILAG-----VT 1694

QY 168 TRG--LPSGICPIV-----VFTGVG-----NVSQGAQEIFKLLPH 201

DB 1695 NFRAYIREGFMEPLLLFPCVGSQFAPYINQIPIIPLSLGLADNDENIRDTAKAGKLIVK 1754

QY 202 TF-----VDAEKLPEIFQAR-----NLKSQSQSTKRVFQYGVVTSRDIIVSHKDPTR- 249

DB 1755 NVATRAVDL-LLPELGRGMFENDRIRLSSVOLTGELLFQVTG--ISSRNEFSEEDGHN 1811

QY 250 -QF-----DKGD-----YYAHPHYTPVFHERIAPYASVIVNCMTWKRFPPL 292

DB 1812 GEFSGLVDVLGDREDRILAAFLVCRNDTSGIVRATTVDIWKALVPNTPRVKEILPTL 1871

QY 293 NMDLOOLMETGCPVGVG-----DITCDIGGSIEFINKSTSIERPFPRYDPSKNSYHDD 347

DB 1872 TGMIVTHLASSSNVLRNIAAQTGLDLVRVGGN-----ALSQLLPSEESLIETSNDS 1925

QY 348 MEGAGVVCLAVDILPTEFKEASQHEFN-----ILSRVLASIASVKQPAELPSYLRACI 402

DB 1926 RQG---VCIALYELIESASTETISQFQSTIVNIRTLALIDESATVREAAALSFQDVV 1982

QY 403 AHAGRLTPLEYIPRMNTM-----IDLAPAKTNPLPKKYSTLV-----SLSGHLDFK 452

DB 1983 GK-----TAVDEVLPYLLHMLESSDNDSDNFALLQLOETMSKSKSDVFFILPTLLAPPIDAF 2038

QY 453 LINEALDIETAGSGPH-----LVRCEVGQSTDDMSYSELEVAGDATTDLKIDUS 503

DB 2039 RASALGSLAEVAGSALYKRLSIINALVDALIGTSEDESTKGALEL-----ALDRVFLS 2092

QY 504 LITSLANEHGDHDAQOETELAKIGKVNETYETDVTIDKGGPKLILGAGRVCPAAEFLEA 563

DB 2093 V-----NDDEGLHPLLOQIMSLKSDNIEK-----RIAVL-----ERLP 2126

QY 564 SYPDICTYGVDDHDADOLHVIVASLYOKDAEETVDGIETNTATOLDVADIGSLDSLSQV 623

DB 2127 NFFDKTVLDFDVIYIPNFVSHALSLDDEED-QRVVNGNFNALSTLLKKVKDKPTLEKLVKA 2185

```

QY 624 EVWISL-----LPASFA-----AIAGV-----CIE 644
Db 2186 KQSLALRGROQDVAARFKLPRGNCVLPFLHGLMYGNSDEREESALADVVSKTPAAN 2245
QY 645 LKKHMTAS-----YVDESMNSLSQAADAGVTIL--CEMGLDPGIDHLS--MKMIDE 694
Db 2246 LKPEVSVITPLIRVVGGERSSDKAAILPALNVLKIPFIPOLQRTFVKLSLD 2305
QY 695 AHARKGKIKAFSTVCGGLPSAANPLAYKFSWNPAGALRGKPNVYKFLGETHVDG 754
Db 2306 ATNETLRLA-AKALGALIEHQPRVDPLVIELVTGAKQATDGVKTAMLKALLEIVMKAG 2364
QY 755 HNYESAKRLRLRELPFALEHLNPNRNSLYDLYGISEASTIYRXYRVEGFSEIMVT 814
Db 2365 SKLNSK-----TNIVNLVEEMLSGNDKLAVY-----AKLIGS 2400
QY 815 LSKTGFDAANHPLLQOTSRPYKGFDELNNISTINTDLDEA-----859
Db 2401 LSEILNDEA-HKILQD-----EVLNADLDGETGKFAILLNSFLKDA 2442
QY 860 -----SGGYDDDLIARLLKGCCKNKEIAVK-----TVTKIKFLGLHEETQIP--902
Db 2443 PTHIENTGLIDFVSVILN-----AIRSPDVYFGENGTTAAGKLLILEGKRSFPV 2493
QY 903 -KGCSSPFDV 911
Db 2494 KKDAAEFPKI 2503

RESULT 11
ID CETP_HUMAN STANDARD; PRT; 493 AA.
AC P11597; Q13987; Q13988;
DC 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cholesteryl ester transfer protein precursor (Lipid transfer protein
DE I).
GN CETP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE-97112972; PubMed-8943225;
RX MEDLINE-97112972; PubMed-3600759;
RA Drayna D., Jarnagin A.S., McLean J., Henzel W., Kohr W., Fielding C.,
RA Lawn R.;
RT "Cloning and sequencing of human cholesteryl ester transfer protein
RT cDNA.";
RL Nature 327:632-634(1987).
RN [2]
RP MEDLINE-97112972; PubMed-2334701;
RX MEDLINE-97112972; PubMed-8943225;
RA Agellon L.B., Quinet E.M., Gillette T.G., Drayna D.T., Brown M.L.,
RA Tall A.R.;
RT "Organization of the human cholesteryl ester transfer protein gene.";
RL Biochemistry 29:1372-1376(1990).
RN [3]
RP MEDLINE-97112972; PubMed-8943225;
RX MEDLINE-97112972; PubMed-8943225;
RA Oliveira C.F.O., Chouinard R.A., Agellon L.B., Bruce C., Ma L.,
RA Walsh A., Breslow J.L., Tall A.R.;
RT "Human cholesteryl ester transfer protein gene proximal promoter
RT contains dietary cholesterol positive responsive elements and mediates
RT expression in small intestine and periphery while predominant liver
RT and spleen expression is controlled by 5'-distal sequences. Cis-acting
RT sequences mapped in transgenic mice.";
RL J. Biol. Chem. 271:31831-31836(1996).
RN [4]
RP MEDLINE-97473500; PubMed-9332354;
RX MEDLINE-97473500; PubMed-9332354;

```

```

RA Williams S., Hayes L., Eisenboss L., Williams A., Andre C.,
RA Abramson R., Thompson J.F., Milos P.M.;
RT "Sequencing of the cholesteryl ester transfer protein 5' regulatory
RL region using artificial transposons.";
RN Gene 197:101-107(1997).
RP [5]
RC SEQUENCE OF 9-493 FROM N.A.
RD TISSUE=Liver;
RA Dinchuk J.E., Hart J.T., Wirak D.O.;
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
RN [6]
RP VARIANT GLY-459.
RX MEDLINE-94013514; PubMed-8408659;
RA Takahashi K., Jiang X.-C., Sakai N., Yamashita S., Bujo H.,
RA Yamazaki H., Kusunoki J., Miura T., Kussie P., Matsuzawa Y., Saito Y.,
RA Tall A.;
RT "A missense mutation in the cholesteryl ester transfer protein gene
RT with possible dominant effects on plasma high density lipoproteins.";
RL J. Clin. Invest. 92:2060-2064(1993)
CC -!- FUNCTION: CETP IS AN EXTREMELY HYDROPHOBIC PROTEIN INVOLVED IN THE
CC TRANSFER OF INSOLUBLE CHOLESTERYL ESTERS IN THE REVERSE TRANSPORT
CC OF CHOLESTEROL.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED MAINLY IN THE LIVER.
CC -!- DISEASE: PROBABLY INVOLVED IN THE DEVELOPMENT OF ATHEROSCLEROSIS.
CC -!- SIMILARITY: BELONGS TO THE BPI/CETP/LBP/PLTP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: M30185; AAA51977.1;
EMBL: M32998; AAA51978.1;
EMBL: M32992; AAA51978.1; JOINED.
EMBL: M32993; AAA51978.1; JOINED.
EMBL: M32994; AAA51978.1; JOINED.
EMBL: M32995; AAA51978.1; JOINED.
EMBL: M32996; AAA51978.1; JOINED.
EMBL: M32997; AAA51978.1; JOINED.
EMBL: U71187; AAD14876.1;
EMBL: AF027656; AAB86604.1;
EMBL: M83573; AAB59388.1;
PIR: A26941; A26941.
MIM: 118470.
DR InterPro: IPR001124; LBP_BPI_CETP.
DR Pfam: PF01273; LBP_BPI_CETP_1.
DR Pfam: PF02886; LBP_BPI_CETP_C; 1.
DR SMART: SM00328; BPI1; 1.
DR SMART: SM00329; BPI2; 1.
DR PROSITE: PS00400; LBP_BPI_CETP; 1.
KW Lipid transport; Cholesterol metabolism; Glycoprotein; Signal;
KW Atherosclerosis; Disease mutation.
FT SIGNAL 1 17
FT CHAIN 18 493 CHOLESTERYL ESTER TRANSFER PROTEIN.
FT CARBOHYD 105 105 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 257 257 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARIANT 459 459 D -> G (IN CETP DEFICIENCY).
FT /FTID=VAR_004172.
FT CONFLICT 251 310 MISSING (IN REF. 2).
FT CONFLICT 422 422 I -> V (IN REF. 2).
SQ SEQUENCE 493 AA; 54770 MW; 16D47ACDC99B063C CRC64;

```

Query Match
Best Local Similarity 19.2%; Score 120; DB 1; Length 493;
Matches 115; Conservative 75; Mismatches 196; Indels 212; Gaps 30;

between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M21041; AAA39490.1; .
PIR; S06467; S06467.
PIR; A40115; A40115.
MGD; MGI:97175; Mcap2.
InterPro; IPR001084; Tubulin-bind.
Pfam; PF00418; tubulin-binding; 3.
PROSITE; PS00229; TAU_MAP; 2.
Microtubules; Repeat; Calmodulin-binding.
DOMAIN 1452 1472 CALMODULIN-BINDING (POTENTIAL).
REPEAT 1662 1692 TAU/MAP MOTIF.
REPEAT 1693 1723 TAU/MAP MOTIF.
REPEAT 1724 1755 TAU/MAP MOTIF.
SEQUENCE 1828 AA; 198980 MW; 200BC59E360538CA CRC64;

Query Match 2.3%; Score 119; DB 1; Length 1828;
Best Local Similarity 20.4%; Pred. No. 19;
Matches 160; Conservative 111; Mismatches 280; Indels 232; Gaps 41;

QY 252 DKGYYAHPHYTPVHERIAPYASVIVNCVMYKRPPLLNMDQLQQLMETGCLVGV 311
DB 741 DEGDYLPPT--TPAVEK-----MPCFPESK-----EEDKAEQAVTG- 778
QY 312 DITCDIGSIEFINKSTIERPFYDPSKNSYHDDMEGAGVCLAVDIILPTFSK 371
DB 779 -----GQTIQV--ETSESSEPF--PAKEYKN-----GTV-MAPD-LPEMLD 813
QY 372 HFGNTLSRLVASLKVOPALPSYLRACIAHAGRLTPLYEYIPRMNTMIDLPAKT 431
DB 814 -----LAGTRSLASVADAEEV---ARRKVSPEAMLAESTSLPPV-----ADES 856
QY 432 PLDPKYSPLVSLGHLFDKFLI---NEALDIETAG--GSFHLVRCVEVGSTD- 480
DB 857 PVTVFQSDLEDGVCVFNKYTVPLSPVQDSNLSGSGSFY-----EGTDKVRD 909
QY 481 ---DMSYSELEVA-----DDTATLDKIIDSLS-----LANEHGDDHAGOEI 527
DB 910 LATDLSLIEVKLAAAGRVKDEFTAEKATPTTSADKSLGSLREFDHRKANDKLT 969
QY 528 -----GKVNNEYETDVTIDKGGPKILILGAGRVCRPAEFASYPDICT 570
DB 970 EEHDSKEHAKSEEMGKVELFGLGITYDQASTKELITTKDTSPKTEKGLSSYPEVA 1028
QY 571 YGVDDHDAQDIHVIVASLYQKDAEIV---DGIE-----NTTATOLD--VADIG 621
DB 1029 -----EVEPTTKADQGLDFAATKAEPQLDKVDFGQM----- 1062
QY 622 QVEWISLLPASPFAATAGVCIELKHHMVTASYVDESMNSLSQAADAGVTILCE 681
DB 1063 -----ASGMNVDAKAIELKFE-----VAQELTSLSEAPQEA- 1102
QY 682 GIDHLSMKMIDEAHARKGKIKAFSTCYGGLPSPAANNPLAYKFSWNPAGALRSKN 741
DB 1103 G--HIKGGKVNTEVKEVKTCP-----DLVHQAADKRESEYE-----SSGE 1145
QY 742 VYKELGETIHVDGHNLYESAKRLRLRELPAFALE-HLPNRSNLIYGDLYGSK 800
DB 1146 LTM---ESLKPDECKKETSPTSILQDEVALKLSVEIPCPVPVSEADL----- 1195
QY 801 ATXRYEGFSEIMVTLSTKGFDAANHPIL-ODTSRPTKYGFLDELLNNISTINT 859
DB 1196 GEVQME-----FIQLPKREESTETPDIPAIPSDVTQPOPEAIVSE---PA 1247
QY 860 SGGYDDDLI-----ARLLKLGCKKNKEIAVKT-----VKTIK--FLGHEET 902
DB 1248 GGEYDKLLFRSDTQISDLLVSESEREFVETCPGELGVVSVVTIEDFTIV-VQT 1306

503 SLTSLANEH-----GGDHAGOEIALALKIGKVEYETDVTIDKGGPKILILG--AGRVCR 556
DB 8 TALLGNNAHACSGKTSHEAG-----IVCRITK-----PALLVLNHHETAKVIQ 49
QY 557 PAAEFELASYPDICTYGVDDHDAQDIHVIVASLYQKDAEETVDGENTTATOLDVADIGSL 616
DB 50 TAFQ--RASYPDI-----TGKAMLLGQVKY-----GLHNIQISHLSTAS----- 88
QY 617 SDLVQSVQVVISLLPASFAHATAGVCIELKHHMVTASYVDESMNSLSQAADK- 673
DB 89 -----SOLVE-----LVEAKSIDVSIQNVSVFRTGLKYGYTT 120
QY 674 LCEMGDLP-----IDHLSMKMIDEAHARKGKIKAFSTCYC-----GGLPSA- 716
DB 121 AWMIGIDQSIDFSDAIDLIQINTQLTCDGSRVRTDAPDCYLSFHKLLHLHQGEREPQWI 180
QY 717 -----RANPLAYKFSWNPAGALRSKNPAVYKELGETI- 750
DB 181 KQLFTNFISFTKLIVLKGQICKEINVISNIMDFVQTRAASILSDGIDGVDISLTDGPVI 240
QY 751 -----HVDGHNLYESAKRLRLRELPAFALEHLPNRSNLIYGDLYGSK 803
DB 241 TASYLESHHKGFHYKNVS-----EDLP-----LPTFSPTLLGD-----SRM 277
QY 804 RYEGFSE-IMVTLSTKGFDAANHPILLODTSRPTKYGFLDELLNNISTINTDLDI--EAS 860
DB 278 LYFWFSERVHSLARVAFQDGR---LMSLGMDEFFKAVLETW-----GFTNQEIFQEVV 329
QY 861 GGYDD--DLIARLLKLG--CCKNKEIAVKTIKTFLGLHEETQIPKGCSSPFDVTCORM 916
DB 330 GGFPSQAQVTVHCLKMPKISCONKGVVNSVVMVKEL-----FPRP 370
QY 917 EORMAYGHN--EODMYLLHHEVEYEP-----DQPAEKHOATLLEFGKVENGRSTT 966
DB 371 DQOHSVAYTFEEDIVT---TVQASYSKKFLSLDLDFTPTVNSNLTSSSESISQSLQ 427
QY 967 ANALFVGPAAAGAL--LLLNKRVQKGV-----IRPQEIYVPALFELISSGKILVE 1018
DB 428 SMITAVGIPEVMSRLEVFTALMNSKGVSLFDIINP-----EITRDGFLLLQ 475

RESULT 12
MAP2_MOUSE
ID MAP2_MOUSE
AC F20357; 1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 2 (MAP 2).
GN MAP2 OR MTAP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=89083571; PubMed=3205744;
RX Wang D., Lewis S.A., Cowan N.J.;
RA "Complete sequence of a cDNA encoding mouse MAP2."
RT Nucleic Acids Res. 16:11369-11370(1988).
RN [2]
RN SEQUENCE FROM N.A.
RP MEDLINE=89043973; PubMed=3142041;
RX Lewis S.A., Wang D., Cowan N.J.;
RA "Microtubule-associated protein MAP2 shares a microtubule binding
RT motif with tau protein."
RN Science 242:936-939(1988).
CC -!- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY
CC STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO
CC SEEM TO HAVE A STIFFENING EFFECT ON MICROTUBULES.
CC -!- SIMILARITY: CONTAINS 3 TAU/MAP REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

QY 903 KGCSSP----FDVICQRMORMAGHNEQDMVLLHHEVEVEYPDGP-----AEKHOATL 953
Db 1307 EGESGSHSVRAAPAPOEERPRPHDEIEIAEAQAQPKDGSFDPATPKEEVAF 1366
QY 954 LEF 956
Db 1367 SEY 1369

RESULT 13
ENO_ZYMO
ID ENO_ZYMO STANDARD; PRT; 428 AA.
AC P33675; 069010;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase) (2-phospho-D-
DE glycerate hydro-lyase).
GN ENO.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93015706; PubMed=1400207;
RA Burnett M.E., Liu J., Conway T.;
RT "Molecular characterization of the Zymomonas mobilis enolase (eno)
RT gene.";
RL J. Bacteriol. 174:6548-6553(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 31821 / ZM4 / CP4;
RA Lee J., Jin S., Kang H.S.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: 2-phospho-D-glycerate -> phosphoenolpyruvate +
CC H(2O).
CC -1- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING
CC THE DIMER.
CC -1- PATHWAY: GLYCOLYSIS.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M99380; AAA27686.1; -.
CC EMBL; AF086791; AAC70360.1; -.
CC PIR; A45732; A45732.
CC HSP; P56252; LPD2.
CC InterPro; IPR000941; Enolase.
CC Pfam; PF00113; enolase; 1.
CC PRINTS; PR00148; ENOLASE.
CC ProDom; PD000902; Enolase; 1.
CC DR PROSITE; PS00164; ENOLASE; 1.
CC KW Lyase; Glycolysis; Magnesium.
CC FT ACT_SITE 155 155
CC FT METAL 242 242
CC FT METAL 287 287
CC FT METAL 313 313
CC FT CONFLICT 186 186
CC FT CONFLICT 369 428
CC FT VETEDTTIADLATVATNGCGIKTSICRSERIAKYNOLMRE
CC EELGSVAKYAGRSVLRKAK -> GGNRRHDC (IN REF.
CC 2).
CC SEQUENCE 428 AA; 45843 MW; 10943357EFC2912 CRC64;

```

```

Query Match 2.2%; Score 118; DB 1; Length 428;
Best Local Similarity 22.6%; Pred. No. 2.5;
Matches 81; Conservative 50; Mismatches 119; Indels 108; Gaps 20;

QY 380 LVASLASVKOPAEPLSYLRACIAHAGRLTPLEYIPRMNTMIDAPARTNPLPKKYS 439
Db 112 LGVSLAVAKAAAE-----ARGLYRYVGG-----TAAHVLVPP---NM 147
QY 440 TLVSLSGHL-----FDKFL-----INEALDIETAGS--FHLVRCV---GQSTD 481
Db 148 NIVNGGMHADNPIDFQEFMIAPVGCASSINEAVRI-----GTEVFHTLKKELSAKMNTNV 202
QY 482 MSYSELEVAGADD-TATLDKIISLTSLANEHGHDHAGOEIELALKIGKRVNEYETDVTID 540
Db 203 GDEGGFAPSLDSASSALDFIVDSISK-----AGYKPGEDVFIALDAASSEFYNDQNIY 256
QY 541 KGGPKILILGAGRVCRPAAEFLASYPDIC---TYGVDD--HD-----ADQIH 582
Db 257 D-----LAGEGRKL-TSAQLVDYVYELCGKYPYISIEDLAEDDFEFWKILTEKLGKRVQ 309
QY 583 VIVASLYQKDAEETVDGIENTTATOLDVA--DIGSLDLVSQVEV--ISLLPASEH--- 635
Db 310 LVGDDLFVTNVKRLSDGIERGIANSLLVKFNQIGSLSETLAAVNMANDASYTPVMSHRV 369
QY 636 -----AALAGVCIELKHMVTAS-----YVDESMSLSQAOKADAGVTIL 674
Db 370 ETEDITIADLATVATNGCGIKTSICRSERIAKYNOLMREIE---ELGSVAKYAGRSVL 424

RESULT 14
ID RPOL_ASFB7 STANDARD; PRT; 1450 AA.
AC P42486;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA-directed RNA polymerase subunit 1 (EC 2.7.7.6).
GN NP1450L.
OS African swine fever virus (strain BA71V) (ASFV).
OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae;
OC African swine fever-like viruses.
OX NCBI_TaxID=10498;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93281390; PubMed=8506138;
RA Yanez R.J., Boursnell M., Nogal M.L., Yuste L., Vinuela E.;
RT "African swine fever virus encodes two genes which share significant
RT homology with the two largest subunits of DNA-dependent RNA
RT polymerases".
RL Nucleic Acids Res. 21:2423-2427(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
RA Rodriguez J.F., Vinuela E.;
RT "Analysis of the complete nucleotide sequence of African swine fever
RT virus".
RL Virology 208:249-278(1995).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC (RNA)(N).
CC -1- SIMILARITY: TO THE CORRESPONDING SUBUNITS OF PROKARYOTIC AND
CC EUKARYOTIC RNA POLYMERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```


Db 1380 SNVLLRMALSSPVQVLTDAAVDSAVNPYIGIAAPTLMGVSVPRIQT 1424

```

CC EMBL; Z21489; CAA79697.1; -
DR EMBL; U18466; AAA65328.1; -
DR InterPro; IPR000722; RNA_pol_A.
DR InterPro; IPR002879; RNA_pol_A2.
DR Pfam; PF00623; RNA_pol_A.1.
DR Pfam; PF01854; RNA_pol_A2.1.
KW Transferase; DNA-directed RNA polymerase; Transcription.
SQ SEQUENCE 1450 AA; 163748 MW; 94D335C50B9A281B CRC64;

Query Match      2.2%; Score 116.5; DB 1; Length 1450;
Best Local Similarity 18.0%; Pred. No. 20;
Matches 159; Conservative 129; Mismatches 306; Indels 291; Gaps 35;

QY 82 PLLDKILEERVSLFDYELIVDDGKSLAFKGFAGRAGLIDFLHGLGQRYLSLIGYSTPFL 141
DB 718 PIMNSINPTNGLF-----QMVATGAKGSPNMIHIMAGIGQIEINTQIQPF 766
QY 142 SLGQSHM-YPSLA-AKAAVIVVAEETATGLPSGICPIVFTVGTGVNVSQGAQEIFKLL 199
DB 767 SFGRLVYYPFALEAQAAYGFCNSYIA-----GLTSPFIF--GEMNGRFDLKNAL 817
QY 200 PHTFVDAEKLPEIFQARNLSKQSTKRVFQLYGCVVTSRDIVSHKDPTRQFDKGDYAH 259
DB 818 STS-----STGYANRAIFGLQSCIV-----DY-- 841
QY 260 PEHYTPVPHERIAPYASVIVNMYWEK-----RFPPLNMQOOLQOLMETGCPVVG 309
DB 842 -----RRVSDITRLVQOVLGDLGDARQLETVRFFETIMLSDO----- 878
QY 310 VCDITCDIGGSTIEFNKSTSTERFF-----RYDPSKNSYHD--DMEGAGVVCVLAVID 361
DB 879 -----ELEDKFKYTQSPLEFEEFRLKKDKDKYQIFLNVENFNSOLLTDV- 927
QY 362 PTEFKEASQHPGNILSRVLASVSKQPAELPSYLRACIAHA-----GR 407
DB 928 -----ROVPNVASIVKNILLSTSGVLFPDEKSLQKVMVTKFCKNLVYVFINNTOER 982
QY 408 L-TPLVEYIPRNRNM-----IDLAKTNPLPDKKYSTVLSGHLFDKFLIN--BALD 459
DB 983 LQTPVPIVYKRAAALMRMLIRIELATVTKTNITCEQMSAIDLRLQVTSQSLINYEAVG 1042
QY 460 ILETAGSGHLYRCEVGQSTDDMSYSELEVAGDDTATLDKIIDSITSLANEHG----- 512
DB 1043 ILAAQVSSEPLTQ-----YMLDSSHRSVAGGTSGVIRVPOEI 1080
QY 513 ---GDHDAQGEIEALKIGKVNEYETDVTIDKGGPKILILGAGRCVCPAAEFELASYPDIC 569
DB 1081 FSAKPVAEQSSMLRL-KNPEVETNKTY-----AQEIANSELI 1120
QY 570 TYGVDDHDAQDIHIVIASL-----YOKDAEEVVDGIENTTATQ--LDVAD----- 612
DB 1121 TF---ERLILQWHLLEYTYSSTKKNVYPDFASDVEMWTDLENHPLLPQPPEDIANWCIR 1177
QY 613 -----TGSLSDIVSQVEVVISLIPASFAHAIAGVIELKHMVYASYVDESMSNLS 663
DB 1178 LELNKTMTMLKSIS-----LESTINSIRA-----KHPNT--YIMHSVENTA 1216
QY 664 QAKADAGVTILCEMLDLP-----GIDHLSMKMIDEAHARKKIKAFKFTSYCGGL 712
DB 1217 -----SGPIIIRIYLRSAFRSTNTRMATDEKIANVVD-----KLLNSTIRGI 1262
QY 713 PSPAANNPLAYKFSWNPAGAL-----RSGKN-----PAYKFLGT 749
DB 1263 PCIRKANVVKLMRURVDAQGLKVLRLDNIYAKTNGTIFGAMLDNDIDPYTVSSSIGDT 1322
QY 750 IHVDGHNLYESAKRLRLRELPAFALEHPN-RNSLYGDLYGISKEASTIYRATXRYEGF 808
DB 1323 MELYG---TEAROKIIEITVTMGDGNHRLHLLYADLMTGTGQVTSLEKAGLNAREP 1379
QY 809 SEIMVTLSTGTFDAANHPLODTSRPTYKGFGLDELNNISTINT 853

```

```

RESULT 15
DYHC_MOUSE STANDARD; PRT; 4644 AA.
AC Q5JHU4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dynein heavy chain, cytosolic (DYHC) (Cytoplasmic dynein heavy chain).
GN DNCHC1 OR DNCH1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB;
RA Sasaki S., Shionoya A., Hirotsune S.;
RL "Complete cDNA sequence of murine cytoplasmic dynein heavy chain.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES.
CC -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
EMBL; AF004877; AAC91078.1; -
DR MGD; MGI:103147; Dnchc1.
KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
FT DOMAIN 48 69
FT DOMAIN 179 200
FT DOMAIN 453 476
FT DOMAIN 541 564
FT DOMAIN 1169 1201
FT DOMAIN 1229 1250
FT DOMAIN 1355 1371
FT DOMAIN 2012 2040
FT DOMAIN 3187 3273
FT DOMAIN 3394 3498
FT DOMAIN 3735 3798
FT NP_BIND 1904 1911
FT NP_BIND 2222 2229
FT NP_BIND 2593 2600
FT NP_BIND 2935 2942
FT NP_BIND 532021 MW; FE5B4E15DD479E1B CRC64;
SQ SEQUENCE 4644 AA; 532021 MW; FE5B4E15DD479E1B CRC64;

```

```

Query Match      2.2%; Score 116.5; DB 1; Length 4644;
Best Local Similarity 19.6%; Pred. No. 1.2e+02;
Matches 214; Conservative 153; Mismatches 406; Indels 321; Gaps 49;

QY 114 FAGRAGLIDFLHGLGQRYLSLIGYSTPFLSGQSHMYPVSLAAAKAAVIVVAEETATF---- 169
DB 2020 YAGRSNLPDLNKLK---FRSLAMTKPDRQLIAQVLMYSQGFRTAEVL--ANKIVPFFKLC 2074
QY 170 --GLPS-----GICPIVFTVGTGVNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKO 221
DB 2075 DEQLSSQSHYDFGLRALKSVLVSAGNVKR--ERIOKIKREKEERGEAVDEGEAENLPEQ 2132
QY 222 SOSTKRVFQLYGCVVTSRDI-----VSHKDPTRQFDKGDYAHPEHYTPVPHERIAPYA 275

```

Db 2133 EILIQSVCEWPKVLAEDIPILFSLSDVFPQVHRGEMTALREELKKVCOEMLTYG 2192
Qy 276 -SVIVNCWYWEKRPPLNMDQQLQMETGCPVLGVCDITCDIGGSEFINKSTSTIERPF 334
Db 2193 DGEVGMWVEK-----VLQYQITQI-NHGLMMVG-----PSGSGKSAWVRVLLKALER-- 2241
Qy 335 FRYDPSKNSYHDMEGAGVVCVLAVIDLPTEFSKE-----ASQHFNGNLSRLV 381
Db 2242 -----LEGVEGVAHIID--PRAISKDHLGYLDPNTRWTDGLFTHVLRKII 2286
Qy 382 ASLASVKQPAELPSYLRACIAHAGRLTPLYEIPMR-----NTMIDLPAKTNPLP-- 434
Db 2287 DNRGELQ-----KRWIVFDGDVDP--EWVENLSVLDNKLTLTPNGERLSLPPN 2336
Qy 435 -----DKYSTLVLSGHLFDKPLINEALDIETAGGSHLVRCVCEVGOSTDDMSYSE 486
Db 2337 VRIMEVQDLKYATLATVS-----RCGMWFSEDEVLSTD 2370
Qy 487 LEVGADDTATLKIIDSLSLSANHG-----GDHDAGOE----- 520
Db 2371 M-----ILNFLARLSRIPLDEGEDEAQRKRKKEGEAEASPMLQIORDAATIMQ 2422
Qy 521 -----IELALKIGKNEYETDVTIDKGGPKILILGA-----GRVCRPAAEFLA 563
Db 2423 PYFTSNGLVTALEHAFLK-----EHIMDLT-----RLRCLGSLFSLHQAACRNVAQYNA 2472
Qy 564 SYPDICTYGVDDHDADQIHVIVASLYQ-----KDAEETVDGIENTTATOLDVADIGSL 616
Db 2473 NHPDPPMOIEQLERYIQRYLYVAILNSLSDSRRLKRAELGEYIRITTVPLTPAPNVI 2532
Qy 617 SDLYSOVEVVISLPPASFAAIGVCIELKKHMVTASYVDESMSNLSQAADAGV----- 671
Db 2533 ID-----YEVISGEWSPQAKVPO--LEVETHKVAAP--DVVYPTLDTVRHEALLYTWLA 2584
Qy 672 ----TILC-----EMGLDPGIDHLSMKMI-----DEAHARKGKIKAFTSYC----- 709
Db 2585 EHKLPLVLGPGSGKWTWLFSAIRALPDMEVGLNFSSATTPPELLKLTEDHYCEYRRTPN 2644
Qy 710 GGLPSA-----AANNPLAYKF-----SW----- 728
Db 2645 GVLAPVOLGKWLVLCFEINLPDMDKYGTQRVISFIQOMVEHAGAFYRTSDOTWVKLRI 2704
Qy 729 -----NPAGALRSKNPAPYKELGE--TIHVDG-----HNLYESAKRLRLRELPAPAL 774
Db 2705 QFVGACNP--PTDPGRKPLSHRFLRPVYVYDYPGPASLTQIYGTFNAMLRLIPSRT 2762
Qy 775 EHLPNRSLYGLDYGISKEASTIYRATXRYEGFSEIMVTLSTKGTGFDAAN-----HP 827
Db 2763 YAEPLTAAMV--EFYTMQOERT--QDTQPHYIYSPRETRWVRGIFEALRPLETLPVEG 2818
Qy 828 LLDQTSRPTYKGFDELNNISTINTDLDIEASGYDDDLIARLKLGCCKNKEIAVKTV 887
Db 2819 LIRIWAHEALRLFQDLRLEDEERRWTDENI-----DMVA--LKHPNIDKEKAMS-- 2866
Qy 888 KTIKFLGLHEETQIPKGCSSPFDVICQRMQORMAYGHNQDMVL-----LHHEVEVEYP 941
Db 2867 RPLYSNNLSKDYIPVDQBELRDYVKRLK--VFYEELDVPVLVNEVLDHVLRIDRI 2923
Qy 942 DQPAEKHQATLLEFGKVENGRSTAMALTVGIPAAICALLLKNKVQTKGVIRPLQPEI 1001
Db 2924 FRQP-----QGHLLIGVSGACKTTLRSRV-----AMNGLSVQIKVHRKYTGDEDEDL 2974
Qy 1002 YVPALEILESSGK 1015
Db 2975 RT-----VLRSGCK 2984

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 15:47:43 ; Search time 41.27 Seconds
(without alignments)
4284.003 Million cell updates/sec

Title: US-09-049-304A-122
Perfect score: 5286
Sequence: 1 CARLLGGGKNGPRVNRIV.....VPALEILSSGKLVKRVET 1022

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTRMBL19.*
- 2: sp_archaea.*
- 3: sp_bacteria.*
- 4: sp_fungi.*
- 5: sp_human.*
- 6: sp_invertebrate.*
- 7: sp_mammal.*
- 8: sp_mhc.*
- 9: sp_organelle.*
- 10: sp_phage.*
- 11: sp_plant.*
- 12: sp_rodent.*
- 13: sp_virus.*
- 14: sp_vertbrate.*
- 15: sp_unclassified.*
- 16: sp_virus.*
- 17: sp_bacteriaph.*
- 18: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5281	99.9	1060	10 004709	004709 zea mays (m
2	3430	64.9	1064	10 004156	004156 arabidopsis
3	3426	64.8	1064	10 004155	004155 arabidopsis
4	3411	64.5	1064	10 004155	004155 arabidopsis
5	3381	64.0	1064	10 004884	004884 arabidopsis
6	1670.5	31.6	465	10 0047M5	0047M5 arabidopsis
7	1481	28.0	482	10 004BT4	004BT4 arabidopsis
8	1466.5	27.7	972	5 09VLX0	09VLX0 drosophila
9	1429.5	27.0	934	5 044503	044503 caenorhabdi
10	1388.5	26.3	926	11 0921I9	0921I9 mus musculu
11	1384.5	26.2	926	11 099K67	099K67 mus musculu
12	1363.5	25.8	926	4 09UDR5	09UDR5 homo sapien
13	1362.5	25.8	926	4 095462	095462 homo sapien
14	746.5	14.1	448	3 09C2G6	09C2G6 neurospora
15	717	13.6	450	3 09PAR4	09PAR4 magnaporthe
16	716	13.5	450	3 059711	059711 schizosacch

17	712.5	13.5	260	10 09FVFB	09Fvfb brassica ol
18	709	13.4	449	3 096TW2	096tw2 penicillium
19	608	11.5	177	10 049176	049176 brassica na
20	602.5	11.4	748	3 096W51	096w51 cryptococcu
21	583	11.0	252	10 09FVFB	09Fvfb brassica na
22	385	7.3	91	10 09FVFB	09Fvfb brassica na
23	317	6.0	386	16 098IX7	098ix7 rhizobium l
24	216.5	4.1	352	17 093312	093312 pyrococcus
25	202	3.8	385	2 09AJC6	09ajc6 bacillus st
26	192.5	3.6	373	17 0978X1	0978x1 thermoplas
27	176.5	3.3	373	17 09HKC0	09hkc0 thermoplas
28	145.5	2.8	1825	11 064715	064715 rattus norv
29	145.5	2.8	1828	11 063724	063724 rattus norv
30	144.5	2.7	367	16 032LL0	032ll0 rhizobium m
31	136.5	2.6	414	16 09KRL3	09krl3 vibrio chol
32	136.5	2.6	1880	12 09DKW2	09dkw2 walrus cali
33	135	2.6	405	16 09RUX8	09rux8 delnocooccu
34	134	2.5	367	16 098E01	098e01 rhizobium l
35	133.5	2.5	846	2 09S6R3	09s6r3 escherichia
36	133.5	2.5	1300	2 032555	032555 escherichia
37	133.5	2.5	1300	2 032591	032591 escherichia
38	133	2.5	792	10 09ARQ9	09arg9 oryza sativ
39	132	2.5	398	16 055131	055131 synecocyst
40	130.5	2.5	2832	2 093N05	093n05 agrobacteri
41	130	2.5	407	17 028990	028990 archaeoglob
42	127.5	2.4	411	17 026953	026953 methanother
43	127.5	2.4	419	16 097RA6	097ra6 streptococc
44	127	2.4	705	16 074535	074535 synecocyst
45	127	2.4	1062	3 09P712	09p712 neurospora

ALIGNMENTS

RESULT	1
004709	PRELIMINARY;
ID 004709	PRT; 1060 AA.
AC 004709;	
DT 01-JUL-1997	(TrEMBLrel. 04, Created)
DT 01-NOV-1999	(TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001	(TrEMBLrel. 19, Last annotation update)
DE	LYSINE-KETOGLUTARATE REDUCTASE/SACCHAROPINE DEHYDROGENASE BIFUNCTIONAL
DE	ENZYME (LYSINE KETOGLUTARATE REDUCTASE/SACCHAROPINE
DE	DEHYDROGENASE).
GN	LKRSDH.
OS	Zeae mays (Maize).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC	Panicoidae; Andropogoneae; Zea.
OX	NCBI_TaxID=4577;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. AGROCIERES F-352 COMMERCIAL HYBRID;
RA	Cord-Neto G., Kemper E.L., Arruda P.;
RT	"Lysine-ketoglutarate reductase/Saccharopine dehydrogenase
RT	bifunctional enzyme from maize seeds.";
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. B73;
RX	MEDLINE=20371120; PubMed=10908876;
RA	Arruda P., Kemper E.L., Papes F., Leite A.;
RT	"Regulation of lysine catabolism in higher plants.";
RL	Trends Plant Sci. 5:324-330(2000).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. B73;
RA	Martinez-Moraes K.C., Kemper E.L., da Silva F.R., Vettore A.L.;
RA	Arruda P.;
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF003551; AAC18622.2;
DR	EMBL; AF271636; ANG21985.1;
DR	SEQUENCE 1060 AA; 116302 MW; 00EB73105F1A4E68 CRC64;

[illegible]

RESULT	3	
004155	PRELIMINARY;	PRT; 1064 AA.
004155;		
01-JUL-1997	(T-EMBLrel. 04, Created)	
01-JUL-1997	(T-EMBLrel. 04, Last sequence update)	
01-DEC-2001	(T-EMBLrel. 19, Last annotation update)	
LYSINE-KETOGLUTARATE REDUCTASE/SACCHAROPINE DEHYDROGENASE BIFUNCTIONAL ENZYME.		
Arabidopsis thaliana (Mouse-ear cress),		
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
eurosid II; Brassicales; Brassicaceae; Arabidopsis.		
NCBI_TaxID=3702;		
[1]		
SEQUENCE FROM N.A.		
STRAIN=LANDSBERG RECTA;		
EMBL=98088001; PubMed=9426595;		
Epelbaum S., McDevitt R., Falco S. C.;		
"Lysine-ketoglutarate reductase and saccharopine dehydrogenase from		
Arabidopsis thaliana: nucleotide sequence and characterization.";		
Plant Mol. Biol. 35:735-748(1997).		
EMBL: U95758; AAB96825.1-;		
SEQUENCE 1064 AA; 117157 MW; 820ACDE332FD1FCB CRC64.		

Distance
 $22.1 \pm 0.0236 \text{ g}$
 0.21 g
 Ground and in $T \leq 10 \text{ m}$

LYSINE-KETOGLUTARATE REDUCTASE/SACCHAROPINE (EC 1.5.1.8).
 F4110.80 OR AT4G33150.
 DE Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Peters S.A., van Staveren M., Dirkse W., Stiekema W.,
 RA Hohseil J., Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
 RA Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RN SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RN SEQUENCE FROM N.A.
 RA Peters S.A., van Staveren M., Dirkse W., Stiekema W., Mewes H.W.,
 RA Lemcke K., Mayer K.F.X.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RL [4]
 RN SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AL035525; CAB36789.1;
 DR EMBL: AL161583; CAB80032.1;
 KW Oxidoreductase.
 SQ SEQUENCE 1064 AA; 117148 MW; 6CA4EBD2898C7E CRC64;

Query Match 64.5%; Score 3411; DB 10; Length 1064;
 Best Local Similarity 63.6%; Pred. No. 4,8e-226;
 Matches 668; Conservative 151; Mismatches 179; Indels 52; Gaps 13:

QY 1 CARLLGGKGNPRVRIIIVQPSRRRIHDAQYEDAGCEISEDLSECGLIIGIKPKQLQM 60
 DB 38 CAR-LLHGGKDRTGISRIIVQVPSAKRIHDAQYEDVGCISEDDLSDCGLILGKPKQPELM 96
 QY 61 ILSDRAYAFFSHTHKAOKENPLLDKILBERVSLFDYELIVGDDGKRSIAFGKFAAGRL 120
 DB 97 ILPERAYAFFSHTHKAOKENPLLDKILSEKVTUCDYELIVGDDGKRLIAFGKFAAGRL 156
 QY 121 IDFLHGLGQRYSLGYSPPFLSGSHMYPFLSAAKAAVIVABEIAITFGLPGICPIVF 180
 DB 157 VDFLHGLGQRYSLGYSPPFLSLGASYMYSSLAATAKAAVIVABEIAITFGLPGICPLVF 216
 QY 181 VFTGVGNYSOGAQEIKLLPHTFVDAEKLPELF-QARNLSKQOSTKRVFQLYGVVTSR 239
 DB 217 VFTGTGNVSLGAQEIFKLLPHTFVPEPSKLPFLFVRKDGISQNGISTKRVYQVYGCIIISQ 276
 QY 240 DIVSHKDPTRFQDGYAHPHYTPVPHERIAPYASVIVNCMYEKRFPPLNMDLQOQ 299
 DB 277 DMVEHKDPKSFQDGYAHPHYTPVPHERIAPYASVIVNCMYEKRFPPLNMDLQOQ 336
 QY 300 LMETGCPYGVGCDITCDIGGSTIEFTNKSTSTERFFRDPSKNSHDDMEGAVVCLAVD 359
 DB 337 LTKKGLPLVIGCDITCDIGGSTIEFTNKRATLIDSPFEFFRNPNNSYDDMDGGLCHAVD 396
 QY 360 ILPTESKEASQHFNGILSRVLASIAKQPAELPSYLRACIAHAGRLTPLYEYIPMR 419
 DB 397 ILPTFAKASQHFNGILSGFVGSIASMTEISDLPAHLKACIASYRGELTSLYEYIPMR 456
 QY 420 NTMIDLAPAKNP-----LPDKKYSTVLSLGHFDKFLINEFALDIETAGS 467
 DB 457 -----KSNPEEAQDNIANGVSSQRTFNILVSLGHLDFKFLINEFALDMIETAGS 507
 QY 468 FHLVRCVGGSTDDMSYSELEVCAADTAIYLDKIDLSLANS-----EHGGDHDAGQEIELA 524
 DB 508 FHLAKCELGSADAESYSELEVCAADRVLDQIIDSILANPNEDYISPHREANKI--S 565
 QY 525 LKIGKNEYETDVTIDKGGPK-----ILLGAGRCVCRPAAEFLASYPDICT-----Y 571
 DB 61 ILSDRAYAFFSHTHKAOKENPLLDKILBERVSLFDYELIVGDDGKRSIAFGKFAAGRL 120

Db. 566 LKIGKVVQ--ENEI---KEKPEMTKKSGVLLILAGRCVCRPAADFLASVRTISSQOQMYKTYF 621
 QY 572 GYVDDHDAQOIHVIVASLYQKDAEETVDCIENTATOLDVADIGSLDLSVQEVVILLSLP 631
 Db 622 GADSEKTDVHVIVASLYLKDAKETVEGIVSDFEAVRLDVSSESLKYSQVDVLSLLP 681
 QY 632 ASFHAAIAGVCIELKKHMVTASYVDESMSNLSSQAADAGVTILICBMGLDGPDIHLSMKM 691
 Db 682 ASCHAVAVKTCIELKHLVTSYVDDETSMLEHAKASAGITILGEMGLDGPDIHLSMKM 741
 QY 692 IDEAHARKGKIKRAFTSYCGGLSPAAANNPPLAYKFSWNPAGALRSKPNVYKFLGETI 751
 Db 742 INDHAIKKGKVSFTSYCGGLSPAAANNPPLAYKFSWNPAGALRSKPNVYKFLGETI 801
 QY 752 VDGHNLYESAKRLRLRELPAFALEHLPNRSNIYGLYGISKEASTIYRATYRYEGFSEI 811
 Db 802 VDGKNLYDSAAFRVFNLPFALECFPNRDSLVYGBHYGIESEATTIFRGTLRYEGFSMI 861
 QY 812 MVTLSKTGFDDAANHPQLQDTSRPTVKYGFDELNNINISTINTDLDIEASGGVDDDLIARL 871
 Db 862 MATLSKLGFPDSEANOVLSGTGRITFGALLSNLTK-----DADNESEPLAG--EEISKRI 916
 QY 872 LKLGCCCKKEIAVKTIKFELGLHEETQIPKGCSSPFDVICORMEQRMAYGHNEQDMVL 931
 Db 917 IKLG--HSKETAAKAAKTIIVFLGFNEEREVPSLCKSVFEDATCYLMEKLAISNEQDMVL 974
 QY 932 LHHEVEVEYDPDQPAEKHQATLLEFGKVENGSRSTTAMALTVGIPAAIGALLLKNKVTOK 991
 Db 975 LHHEVEVEFLESKRIEKHTATLLEFGDKIKNGQTTAMAKTVGIPAAIGALLIEDIKIR 1034
 QY 992 GVIRPQPEIYVPALEILESSGKIKVEKE 1021
 Db 1035 GVLRLPEAEVILPALDILQAYGIKLMEKAE 1064

RESULT 5
 ID 004884 PRELIMINARY; PRT; 1064 AA.
 AC 004884;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE LYSINE-KETOGLUTARATE REDUCTASE/SACCHAROPINE DEHYDROGENASE (EC 1.5.1.8)
 DE (SACCHAROPINE DEHYDROGENASE (NADP+, L-LYSINE FORMING)) (EC 1.5.1.9).
 DE LKR/SDH.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLEMBIA; TISSUE=SEEDLING HYPOCOTYLS;
 RA Tang G., Miron D., Zhu-Shimoni J.X., Galili G.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: N6-(L-1,3-DICARBOXYPROPYL)-L-LYSINE + NADPH.
 CC + H(2)O = L-LYSINE + 2-OXOGLUTARATE + NADP.
 DR EMBL: U90522; AAB53975.1;
 DR EMBL: U90523; AAD00700.1;
 KW Oxidoreductase.
 SQ SEQUENCE 1064 AA; 117023 MW; 3DD075F9520ED7A9 CRC64;

Query Match 64.0%; Score 3381; DB 10; Length 1064;
 Best Local Similarity 63.3%; Pred. No. 5,6e-224;
 Matches 665; Conservative 150; Mismatches 183; Indels 52; Gaps 13:

QY 1 CARLLGGKGNPRVRIIIVQPSRRRIHDAQYEDAGCEISEDLSECGLIIGIKPKQLQM 60
 Db 38 CAR-LLHGGKDRTGISRIIVQVPSAKRIHDAQYEDVGCISEDDLSDCGLILGKPKQPELM 96
 QY 61 ILSDRAYAFFSHTHKAOKENPLLDKILBERVSLFDYELIVGDDGKRSIAFGKFAAGRL 120


```
Db 97 ILPERAYAFFSHTHKAQENMPLLDKILSERVTLCDYELIVDGHGKRLLAFFGKYAGRAGL 156
Qy 121 IDFLHGLGORYLSLGYSTPFLSLGSHMYPSLAATAKAIVVVAEIAETFGLPSCICPIVF 180
Db 157 VDFLHGLGORYLSLGYSTPFLSLGASYMSYSSLAATAKAIVSVGEISAGQGLPGICPLVF 216
Qy 181 VFTGVGNVSGQAQEIFKLLPHTFVDAEKLPFIF-QARNLSKQSQSTKRVFQYLCVWTSR 239
Db 217 VFTGTGNVSLGAQEIFKLLPHTFVDAEKLPFIF-QARNLSKQSQSTKRVFQYLCVWTSR 276
Qy 240 DIVSHKOPTROFDGXYAHPEHYTPVFERIAPYASVIVNCWMEKFPFLLNMQLOQ 299
Db 277 DMVEHKDFSKFDRADYIYAHPEHYTPVFERIAPYASVIVNCWMEKFPFLLNMQLOQ 336
Qy 300 LMETGCPVLGVCDITCDIGGSIEFINKSTIERPFYDPKSNVHDDMEGAGVYCLAVD 359
Db 337 LTKKGLPLVGICDITCDIGGSIEFINKSTIERPFYDPKSNVHDDMEGAGVYCLAVD 396
Qy 360 ILPTFEKSAQSHFGNLTSLRSLASVQPAELPSYLRACIAHAGRLTPLYEYIPMR 419
Db 397 ILPTFEKSAQSHFGNLTSLRSLASVQPAELPSYLRACIAHAGRLTPLYEYIPMR 456
Qy 420 NTMIDLAPAKTNP 432
Db 457 -----KSNP 460
Qy 468 FHLVRCVGGQSTDDMSYSELEVAGDATTATLDKIISLTSLSAN---EHGGDHDAGQETELA 524
Db 508 FHLAKCELGSADAESYSELEVAGDADRVLDQIISLTLRANPNEDYISPHREANKI--S 565
Qy 525 LKIKVNEYETDVIDGGPK-----LDPDKYSTLVLSLGHFLFKFELINEALDIETAGGS 467
Db 566 LKIKVQVOO-ENEI---KEKPEMTKSGVLILGAGVRCRPAADFSLASVRTISSQOWMYKTF 621
Qy 572 GVDHDDADQIHVIIVASLYQKDAETVDGIENTTATQLDVADIGLSLDSVQVWISLLP 631
Db 622 GADSEKTDHVIIVASLYLKDQETVSGISDVEARLDVDSSELLKLYVSQVDVLSLLP 681
Qy 632 ASFHAAIAGVCIEUKHMVTASYVDESMSNLQAAKADAGVTILCEMGIDPCIDHLSMKM 691
Db 682 ASCHAVYAKTCIELKHLVTSYVDDETSMLENKASAGITILGEMGLDPCIDPMMAKM 741
Qy 692 IDEAHARKGKAKTSYCGGLPSAAANNPLAYKFSNPNAGALRSGKNPAAVYKPLGETIH 751
Db 742 INDAPITKGVKFTSYCGGLPSAAANNPLAYKFSNPNAGALRSGKNPAAVYKPLGETIH 801
Qy 752 VDGHNLYESAKRLRLRELPALFALEHLPNRLSLIYGLYISKEASTYRATXRYEGFSEI 811
Db 802 VDGNLYDSAAFRVPLNPAFALECPNRLSLYGEHYGIESATYIFRGTLRFGFSMI 861
Qy 812 MVTLSTGTFDAAHNPQLQDTSRPTYKGFLELLNNTISTINTDIDTASGGYDDDLIARL 871
Db 862 MATLSKLGFFDSEANOVLSGKRITFGALLSNILNK---DADNESEPLAG-EEISKRI 916
Qy 872 LKLGCCNKKEIAVKTITKFLGHEETQIPKCCSPFDVICORMEORMAYGNEQDMVL 931
Db 917 IKLG--HSEKTAKAAKTIVFLGNEEREVPVSLCKSVFDTATCYLMEELKAYSNEQDMVL 974
Qy 932 LHHEVEVEYPDGQAPAKHOATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKKNVOTK 991
Db 975 LHHEVEVEFLSKRIEKHTATLLEFGDKINGQTTTAMAKTVGIPAAIGALLLLIEDKIKTR 1034
Qy 992 GVTPLQPEYVPALEILESSGKIKVEKE 1021
Db 1035 GVLRLPEAEVYLPALDILQAYIKLMEKE 1064
RESULT 6
Q947M5 PRELIMINARY; PRT: 465 AA.
AC Q947M5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DB 01-DEC-2001 (Tremblrel. 19, Last sequence update)
```

```
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MONOFUNCTIONAL LYSINE-KETOGLUTARATE REDUCTASE.
GN LKR.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Tang G., Zhu X., Galili G.;
RT "Arabidopsis monofunctional lysine-ketoglutarate reductase.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF95389; AAK97099.1; -.
SQ SEQUENCE 465 AA; 51498 MW; CE27C04A2F9C12ED CRC64;

Query Match 31.6%; Score 1670.5; DB 10; Length 465;
Best Local Similarity 71.8%; Pred. No. 1e-106;
Matches 311; Conservative 59; Mismatches 52; Indels 11; Gaps 3;

Qy 1 CARLLGGGKNGPRVNRRIIVQPTTRRIHDAQYEDAGCEISEDLSEGLIIGIKOPKLOM 60
Db 38 CAR-LLHGGKDRGTGISRVVQPSAKRIHHDALYEDVGEISDDLSDCGLILGKQPELEM 96
Qy 61 ILSDRAYAFTSHTHKAQENMPLLDKILSERVTLCDYELIVDGHGKRLLAFFGKYAGRAGL 120
Db 97 ILPERAYAFFSHTHKAQENMPLLDKILSERVTLCDYELIVDGHGKRLLAFFGKYAGRAGL 156
Qy 121 IDFLHGLGORYLSLGYSTPFLSLGSHMYPSLAATAKAIVVVAEIAETFGLPSCICPIVF 180
Db 157 VDFLHGLGORYLSLGYSTPFLSLGASYMSYSSLAATAKAIVSVGEISAGQGLPGICPLVF 216
Qy 181 VFTGVGNVSGQAQEIFKLLPHTFVDAEKLPFIF-QARNLSKQSQSTKRVFQYLCVWTSR 239
Db 217 VFTGTGNVSLGAQEIFKLLPHTFVDAEKLPFIF-QARNLSKQSQSTKRVFQYLCVWTSR 276
Qy 240 DIVSHKOPTROFDGXYAHPEHYTPVFERIAPYASVIVNCWMEKFPFLLNMQLOQ 299
Db 277 DMVEHKDFSKFDRADYIYAHPEHYTPVFERIAPYASVIVNCWMEKFPFLLNMQLOQ 336
Qy 300 LMETGCPVLGVCDITCDIGGSIEFINKSTIERPFYDPKSNVHDDMEGAGVYCLAVD 359
Db 337 LTKKGLPLVGICDITCDIGGSIEFINKSTIERPFYDPKSNVHDDMEGAGVYCLAVD 396
Qy 360 ILPTFEKSAQSHFGNLTSLRSLASVQPAELPSYLRACIAHAGRLTPLYEYIPMR 419
Db 397 ILPTFEKSAQSHFGNLTSLRSLASVQPAELPSYLRACIAHAGRLTPLYEYIPMR 456
Qy 420 NTMIDLAPAKTNP 432
Db 457 -----KSNP 460

RESULT 7
Q94BT4 PRELIMINARY; PRT: 482 AA.
AC Q94BT4;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DB 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE AT4G3150/F4I10_80.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Gibson H.A., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
```

RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yu G., Yu S., Shinozaki K.,
RA Davis R.W., Theologis A., Ecker J.R.;
RA "Arabidopsis cDNA clones";
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY039906; AAK64010.1; -
SQ SEQUENCE 482 AA; 52655 MW; 5AB85A044BD7B942 CRC64;

Query Match 28.0%; Score 1481; DB 10; Length 482;
Best Local Similarity 60.7%; Pred. NO. 1.2e-93;
Matches 293; Conservative 75; Mismatches 101; Indels 14; Gaps 5;

QY 546 IILGAGRCRPAAFSLASYPDIC-----TV-GVDDHDAQIHVIVASLYOKDAETVD 598
DB 7 VILGAGRCRPAAFSLASVRTISSQWKTYTGVADSEKTDVHVIVASLYLKDAKEIVE 66
QY 599 GIENTTATOLDVADIGSLSDLYSQVEVWISLLPASFAHAGVCIELKKHMYTASYVDES 658
DB 67 GISDVEAVRLVDSSESLLKYQSDVVLSPASCHAVVAKTCIELKKHLVTASYVDDE 126
QY 659 MNLNSOAKDAGVTILCEMGIDPGIDHLSMKMIDEAHARKKIKAFTSYCGGLSPAAA 718
DB 127 TSNLHEKAKSAGTITLGENGLDPGIDHMMAMKMINDAHIKKGVKVSFTSYCGGLSPAAA 186
QY 719 NNPLAYKFSNPAAGLRSGNPAVYKFLGETIHVDGHNLYESAKRLRLRELPFALEHLP 778
DB 187 NNPLAYKFSNPAAGLRSGNPAVYKFLGETIHVDGHNLYESAKRLRLRELPFALEHLP 246
QY 779 NRNSLYGLYGISKEASTIYRATRYEGFSEIMVTLSTKGTGFFDANHPHLLQDTSRPTK 838
DB 247 NRDSLYGHEGYSEATTIFRGLTRYEGFSEIMATLSKLGFFDSEANQVLTGKRTIFG 306
QY 839 GFDELNNISINTDLDTEASGVDDDLIARLLGCKCKNEIAVKTIKFLGLHDEE 898
DB 307 ALLSNLTK-----DADNESEPLAG-EEESKRLIKLG--HSKTRAKAAKTIVFGFNEE 359
QY 899 TQTPKGCSPFDVICORMEQRMAYGHNEQDMVLLHHEVEYEPDGPQAEKHQATLLEFGK 958
DB 360 REVPSLCKSVFDTATCYLMEELKAYSQNEQDMVLLHHEVEYEPDGPQAEKHQATLLEFGD 419
QY 959 VENGRTTAMALTVGPIPAAGALLLNKVKQVTRPLQPELYVPALLESSTGKLVE 1018
DB 420 IKNGQTTAMAKTVGPIPAAGALLLNKVKQVTRPLQPELYVPALLESSTGKLVE 479
QY 1019 KVE 1021
DB 480 KAE 482

RESULT 8
Q9VLX0 PRELIMINARY; PRT; 972 AA.
AC Q9VLX0
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG7144 PROTEIN.
GN BEST:CK02318 OR CG7144.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Cocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Fostier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; -
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003618; AAF52559.1; -
DR FlyBase: FBgn0025687; BEST:CK02318.
DR FlyBase: FBgn0030715; Or13a.
SQ SEQUENCE 972 AA; 108214 MW; 7DA8002E16DBC1E9 CRC64;

Query Match 27.7%; Score 1466.5; DB 5; Length 972;
Best Local Similarity 33.4%; Pred. No. 3.8e-92;
Matches 343; Conservative 183; Mismatches 318; Indels 183; Gaps 22;

QY 17 RIIVQPSRIHHDQAYEDAGCEISEDLSECGLIIGIKPKQLMILSDRAYAFESHKTA 76
DB 98 KVIVQPSNRRAVPMQAYMOAGAHQIEDISDAVIGVKQVPIDALIPGKTYCFESHKTA 157
QY 77 OKENMPLDKILEERVSPDYELIVGDGKRSIAFGKFAGRAGLIDFLHGLQVLSLY 136
DB 158 QESNMPLDLAILEKKIRLDYERIDERGARQVAFGKYAGVAGMVLHGLRLALGH 217
QY 137 STPFSLGQSHMYPSLAAKAAVIVVAEIAATFGLPSGICPIVFTGVGNVSOQAEIF 196
DB 218 HTPFMHIGPAHYRNSSMARQAIRDCGYEISLGMKMPKSGITFTFTGSGNSQGAQYF 277
QY 197 KILPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDVSVDKDPTRQFDKGY 256
DB 278 SELPIEYVPEMLRKVAEHCNQK-----LYGCEVSRSDHLERRE-GGGFQKEY 326
QY 257 YAHPEHYTFVHERIAPYASVIVNCMYWBEKFRPPLNMDLOQLM-----ETGCP 306
DB 327 DEFFERYISTFTSTKIAPYASVIVNGIYVAVGSPKLSIPDAKNLLRPANTPWLPTSGSP 386
QY 307 -----LVGVCDITCDIGGSIETINKSTSERFFRDPSKNSVHDDMEGAGVCLAVDIL 361
DB 387 ALPHRLMIAICDISADPGSGIETFMNCTTIDTPFCYLDADRNDKTKSKFGVGLVCSIDNM 446
QY 362 PTEFSKEASOHFGNLSRLVASLAVSKQPAELP-----SY-LRRACIAHAGRLTYEY 414
DB 447 PQLPRESTDGLGELLAPHVHDI--IKSDAKKPLAEENFSPYQSAIINSGQLTEGFQY 504
QY 415 IPRMRNTMIDLAPAK7NPLPDKKYSTLVSLSGHLFDKFLINEALDIETAGGSFHLVRC 474

Db 505 IQELR----- 509
QY 475 VQSTDDMSYSELEVAGADDTATLQKIDSLSLANEGHGDHAGQETELALKIGKVNEYE 534
Db 510 --ESQSHRSRHKME--GSSE----- 526
QY 535 TDVTIDKGGPKILILGAGRVCRAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAE 594
Db 527 -----DK--KVLVLGAGMVSAPLVEWL-----HREKDVSIIVCSOVKEAD 565
QY 595 ETVDGIENTTATQDVA-DIGSLDLSVQVEVWISLIPASFHAAIAGVCIELKKHMYTAS 653
Db 566 RLAAQYAGVDSVYLDVNESTGHLOELGGRADVYVSVLLPSYSLHGMVARYCVAEGTHMYTAS 625
QY 654 YVDESMNSLSQAADKAGVTILCEMGLDGGIDHLSMKMKIDEAHARKKIKAFTSYCGGLP 713
Db 626 YLNDEISGLHEEAKAGVYTNNEVGLDGGIDHLLALECIIHEVQDKGAVVESFVSYCGGLP 685
QY 714 SPAANNPLAYKFSWNPAGALRSCKNPAYVKFGLGETIHDVGHNLYESAKRLRLRELPAFA 773
Db 686 APEHNNALAYKFSWNPAGALRSCKNPAYVKFGLGETIHDVGHNLYESAKRLRLRELPAFA 744
QY 774 LEHLPNNSLIYDGLYISKEASTIYRATRYEGFSIMVTLTKTFFDAANHPLQDTS 833
Db 745 LEGFPNRDSTKYGNLYGLGRDVHTLLRGTIRYKGFSESIKPMQLLGLIDPEPHALLHPSG 804
QY 834 RP-TYKGFDELIN-NISTINTDLDIEASGYDDDLTARLLKLGCKCKEIAVTKYTKIK 891
Db 805 PDVTRQVLVHLMGMSDSTI-----FYENLQKRLTE-----RIGDVGIE 844
QY 892 FLGLHEETQPKGCSFPDVCQRMQRMAYGHNEQDMVLLHHEVEVEYPDGQPAKHQA 951
Db 845 SLGLDDTTPYVK-LNTPDLTSLYLSKRLAFERDRLVLRREVGLRWPDGR-REERGI 902
QY 952 TLEFGKVENGRSTTAMALTVGTPAAGALLLLKKNVQKGVTRPLOPEIYVPALEILES 1011
Db 903 NFVYVQGPQ-GHS--AMAMVFGPAALAAKMLDGEIQRGVLLPFTPDYRMLQLRS 959
QY 1012 SGTKLVE 1018
Db 960 EGLTATE 966

RESULT 9
O44503
ID O44503 PRELIMINARY; PRT; 934 AA.
AC O44503;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE R0203.1 PROTEIN.
GN R0203.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M., Berts M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du X., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Keshaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans."

Nature 368:32-38(1994).
[2]
RL SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RA Wu X., Antoniou B.,
RT "The sequence of C. elegans cosmid R0203.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RC Waterston R.,
RA EMBL: A9034615; AAB94142.1; -;
DR EMBL: A9034615; AAB94142.1; -;
SQ SEQUENCE 934 AA; 103751 MW; CF985E127DDA69D5 CRC64;

Query Match 27.0%; Score 1429.5; DB 5; Length 934;
Best Local Similarity 32.5%; Pred. No. 1.2e-89;
Matches 334; Conservative 190; Mismatches 321; Indels 183; Gaps 24;
QY 10 KNGPRVNRRIIVOPSTRRIHDAQYEDAGCEISDLSEGLIIGIKPKLQMLSDRAYAF 69
Db 58 KSG--VN-VLIQPSNRRAYPIQDYISAGATVREDLSEAHILMSKVSVPIDQLIPNKTYAF 114
QY 70 FSHTHKAOKENMPLLDKILBEERSVSLFDYELIVGDDGKRSALAFKGFAGRLIDFLHGLQ 129
Db 115 FSHTHKAOKENMPLLDKILBEERSVSLFDYELIVGDDGKRSALAFKGFAGRLIDFLHGLQ 129
QY 130 RYLSLGYSTPFLSLGSGHMYPSLAAKAAVIVVAEETATFGLPSPGICPIVFTGVGNVS 189
Db 175 RLLALGHNTFFLHGLAHNYNSHMAINALRDIGYELALDRMPRSGLPIFVFTSGNVS 234
QY 190 OGAQETFKLLPHTFVDAEKLPEIFQARNLSKQSQTQKRVFQYGVVTSRD--IVSHKDP 247
Db 235 OGARELFELHPHEVYDVATLPKVAQKQLNK-----VYGVVTRKDHVPKHGCP 284
QY 248 TRQFDKGYAHPEHYTPVFEHERIAPYASVIVNCMYWEKARFPPLNMDLOQLM----- 301
Db 285 ---FKKEFEQPDRTYSKFATEIAPYASVIVNCMYWEKARFPPLNMDLOQLM----- 301
QY 302 -ETGCP-----LVGVCDITCDIGSIEFINKSTIERPFYFRYDPSKNSVHDDMEGA-GVV 354
Db 342 DTPGCGPTLPHLLALCDISADPGSGVEFMRECTIDKFAIYDADFNTSSDSDFAASCL 401
QY 355 CLAVDILTEFSKASQHFNGNLSRLVASLASKVP-----AELPSYLRACIAHAGRLT 409
Db 402 VCSIDNMPAQMPIEATEQFNGNLLYPLWLDLNTSDQHFDRLCQRTKNAITDQGLIT 461
QY 410 PLYEYIPMRNTMIDLAPAKTNPLDPKKYSTLVSLGHLFDKFLINEALDIETAGGSFH 469
Db 462 PNFYIAQLRKDKASASANSRVM----- 485
QY 470 LVRCEVGSTDDMSYSELEVAGADDTATLQKIDSLSLANEGHGDHAGQETELALKIGK 529
Db 486 -----GG----- 487
QY 530 VNEYETDVTIDKGGPKILILGAGRVCRAAEFLASYPDICTYGVDDHDADQIHVIVASLY 589
Db 488 -----TTDK-----VLLTGAGMVGSPFADYFKQKD-----VNLTVATES 523
QY 590 QKDAEETVDGIENTTATQDVA-DIGSLDLSVQVEVWISLIPASFHAAIAGVCIELKKH 648
Db 524 QRDGRLCTS-PNIOQSVVVDIARESHMTMERLIREHDLVSVLLPFPNPLVAKMISNQRD 582
QY 649 MYTASYVDESMSLSQAADKAGVTILCEMGLDGGIDHLSMKMKIDEAHARKKIKAFTSY 708
Db 583 MYTSSYVSEPEALDKAAKADADVIMNEAGLDGIDHLMAMECFDDIKEHGGRITSPEF 642
QY 709 CGGLPAPAEWSNPPLRYKFSWNPAGALRSCKNPAYVKFGLGETIHDVGHNLYESAKRLRL 768
Db 643 CGGLPAPAEWSNPPLRYKFSWNPAGALRSCKNPAYVKFGLGETIHDVGHNLYESAKRLRL 768
QY 769 LPFALEHLNPNRNSLIYDGLYISKEASTIYRATRYEGFSIMVTLTKTFFDAANHPL 828

Db	701	MPGLNIGPNDKSTKSYDVGGLGNDCKTIIIRGTRYQGFVDTVKALHSVGLSGDNIDS	760
Qy	829	LQDTSRP--TYKGFDELNNISNTINTDIDIEASGGYDDDLIARLLKLGCKCKKEIAVKT	886
Db	761	FTSNIGPDLTWK-----ELIASLS--NOKLDI-----FPDSLRHII-----EERVGKG	802
Qy	887	VKTIKFLGHEETQIPKGCSSPFDVICQRMQEMAYGHNEQDVLHHEVEYPDGQPA	946
Db	803	LSALENLGLFSKVDVRH--GTPIDTIAQYLAKILAFKDHESDLVNLNHDIGAQLPQGN--S	860
Qy	947	EKHOATLLEFGKVENGRSTTAMALTGIPAAIGALILKKNVQTKGVIRPLOPEIYVPAL	1006
Db	861	ERRHISLVQIG---NPNGFSAMARTVGYTTAIVSHVNLNNEIORAGIQRPIILKEIVRPAL	917
Qy	1007	EILESSGI 1014	
Db	918	KRLRDFGI 925	
RESULT	10		
ID	Q921I9	PRELIMINARY; PRT; 926 AA.	
AC	Q921I9;		
DT	01-MAY-1999 (TEMBLrel. 10, Created)		
DT	01-MAY-1999 (TEMBLrel. 10, Last sequence update)		
DT	01-DEC-2001 (TEMBLrel. 19, Last annotation update)		
DE	LYSINE KETOGLUTARATE REDUCTASE/SACCHAROPINE DEHYDROGENASE.		
GN	LORSOH OR LKR/SDH.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Strausberg R.;		
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: BC005420; AA05420.1;		
DR	MGI:1353573; LORSDH.		
DR	InterPro: IPR002016; Peroxidase.		
SQ	SEQUENCE 926 AA; 102946 MW; E01118983B27A2B5 CRC64;		
Query Match	26.3%;	Score 1388.5;	DB 11; Length 926;
Best Local Similarity	31.7%;	Pred. No. 8.2e-87;	
Matches	327;	Conservative 186;	Mismatches 316; Indels 201; Gaps 22;
Qy	17	RIIVQSTRRTHDAQYEDACEISEDLSPGGLIIGIKOPKLOMILSDRAYAFESHTHKA	76
Db	56	KVLIQPSNRRAHKEYVRAGQIQEEDTEACILGWKRPPEEKMLSKYIYAFESHTHKA	115
Qy	77	OKENPLDKLIERVSLFDYELIVGDDGKRSIAFGKFAGRAGLIDFLHGLGQRYLSLG	136
Db	116	QANNLLDEVLKQELIDYKWDHGRSRIYAFAGWAGVAGMINILHGMGLRLLALGH	175
Qy	137	STPFLSLGOSHWYPSLAARVIVAEIATFGLPSCICIVFVFGVGNVSGOAEIF	196
Db	176	HTPFHGLGMAHNYRNSOAVQAIADAGYEISLGLMPKSIGLTFVFTGTGNVSKGAQEVF	235
Qy	197	KLLPHTFVDAEKLPIFOARNLKSQSTKRVFOLYGCVTSRDVSFKDPTQ-----FD	252
Db	236	NELPEYVEPHELREVSXTGLRK-----VYGVLS-----RHHHLVRKTDGYD	280
Qy	253	KGDYYAHPEHTYTPVFERIAPSYVIVNCMWYKRFPLLNMDQLQQLM-----ET	303
Db	281	PVEYKYPERTSRENTDIAPYTTCLINGIYWEQNTPRLLTRQDAQSLLVFKSSVWPVE	340

Qy	304	GCP-----LVGCDITCDIGSIEFINKSTSIERPFYRDPKSNYSHDDMEGAGVVCVAY	358
Db	341	GCPELPHKLVAICDISADTGSGIDFMTCTTIERPFCMDAQOIIHDSVEGSGILMCSI	400
Qy	359	DILTFEFSKASQHFNGNLSRLVAS--LASVKQPAELPSY--LRRACIAHAGRLTPLYE	413
Db	401	DNLPAQLPIEATYFEGDMLYPYVEEMLLSDASQPLESONFSPVVRDAVITNSGLLTDKYK	460
Qy	414	YIPMRNTMIDLAPAKTNPLPKKYSTLVSLSGHLFDKFLINEALDIITETAGGSFHLVRC	473
Db	461	YIOLRES-----	468
Qy	474	EVGQSTDDMSYSELEVAGADDTATLKIIDSLSLANEHGGHDAGQETELAKIKGVNEY	533
Db	469	-----RERTQFLSMSTK-----	480
Qy	534	ETDVTIDKGGPKILILGAGVRCRPAAEFLASYPDI--CTYGVDDHDADQIHHIVIASLYQKD	592
Db	481	-----KKVLVIGSGYVGPVLEYLSRDNNITEITLGS-----MTNQMQQLS	521
Qy	593	ABETVDGIENTTATQDVADIGSLSDLVSVQVEVITSLPASFHAAIAGVCIELKKHMTA	652
Db	522	KKYNINPVSLTVGKQ-----EAKQLSVESQDLVLSLLPYVLHPYVAKACIESRYNMVTA	576
Qy	653	SYVDESMSNLSSAAKADAGVTILCEMGLDPGIDHLSMKMIDEAHARKGKIKAFSTYCGGL	712
Db	577	SVITPAMKELEKSVDGAGITVIGELGDPGLDHMLAMETIDTAKELGATVESYVSYCGGL	636
Qy	713	PSPAANNPLAYKFSWNPAGALRSKGNPAVYFGLGETIHDVGNHLYESAKRLRLRELPAF	772
Db	637	PAPEHSDNPLRYKFSWSPVGLNMIMQPSYLLNGKVVNVTVGVSFLNSV--TPMDYFPGL	695
Qy	773	ALEHLPNRNSLYGDLYGSKSEASTIYRATYRXYEGESEMVTLSKTGFFDAAHHLLQDT	832
Db	696	NLEGYPNRDSIKYAEIYGIS--SAHTLLRGTRYKGYSKALNGFVKGLINREAYPALRPE	754
Qy	833	SRP--TVYKGFDELNNISNTINTDLDIEASGGYDDDLIARLLKLGCKCKKEIAV-----	884
Db	755	ANPLTWKQLCDDLVG-----ISR--SSPCKEKEVVFVTKLGGDN	791
Qy	885	KTVKTIKFLGLHBEETQIPKGCSSPFDVICORMEQRNAYGHNEQDVLHHEVEYEPDQ	944
Db	792	TQLEAAEWLGLLGDQVPO--AESIVDAFSKHLVSKLSYSGYPEEKDMIVMRDSFGIRHPSGH	850
Qy	945	PAEKHOATLLEFGKVENGRSTTAMALTGIPAAIGALILKKNVQTKGVIRPLOPEIYVP	1004
Db	851	AENKTTIDLIVYGDF--NGFS--ANAKTVGLPTAAAKMLLDGEIEAKGLMGFTKEIYGP	906
Qy	1005	ALEILESSGI 1014	
Db	907	ILERIKAEGI 916	
RESULT	11		
ID	Q99K67	PRELIMINARY; PRT; 926 AA.	
AC	Q99K67;		
DT	01-JUN-2001 (TEMBLrel. 17, Created)		
DT	01-JUN-2001 (TEMBLrel. 17, Last sequence update)		
DT	01-DEC-2001 (TEMBLrel. 19, Last annotation update)		
DE	LYSINE OXOGLUTARATE REDUCTASE, SACCHAROPINE DEHYDROGENASE.		
GN	LORSOH.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Strausberg R.;		
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: BC005420; AA05420.1;		
DR	MGI:1353573; LORSOH.		
DR	InterPro: IPR002016; Peroxidase.		

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ007619.2; -
DR InterPro: IPR002016; Peroxidase.
DR PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1.
SQ SEQUENCE 926 AA; 102147 MW; DBFD9E17EF5CB680 CRC64;

Query Match 25.8%; Score 1362.5; DB 4; Length 926;
Best Local Similarity 32.2%; Pred. No. 5.1e-85;
Matches 326; Conservative 174; Mismatches 337; Indels 181; Gaps 20;

QY 257 YAHPEHYTPVHERIAPYASVIVNCMYWEKRPPLINMDQLQOLME-----TGCP- 306
DB 285 DKHPERYISRFTDIAPYTCCLINGIYWEONTPLRLTRODAQSLAPGKFSFAGVCGCPA 344
QY 307 ----LVGCDITCDIGGSIEFINKSTSTERPFYRDPKSNVHDDMEGAGVVCVLAVIDLP 362
DB 345 LPHKLVAIICDISADTGGSTIEFTECTTIEHPFCMDADQHIITHDSVEGSLMCSIDNLP 404
QY 363 TFSKASQHFNGILSRVAS--LASVKOPAEPLSY--LRRACIAHAGRLTPLYEYIPR 417
DB 405 AQPLIEATECFGDMLYPYVEEMILSDATPLESQNPFVVRDAVITNGTLDPKYYIQT 464
QY 418 MRNTMIDLAPAKTNPLPDKKYSTLVSLGHLFDKFLINEALDIIETAGGSFHLVRCEVQ 477
DB 465 LRES----- 468
QY 478 STDMSYSELEVGADDTATLTKIISLTSLANEHGGHDHAGOEIELALGKGVNEYETDV 537
DB 469 -----RERAQSLSMGTR----- 480
QY 538 TIDKGGPKILILGAGRVCRPAAEFLASYPDI-CTYGVDDHDADQIHVIIVASLYQKDAEET 596
DB 481 -----RKVLVLSGYISEPVLEYSRDNIEITVGS-----MKNQIEQ 519
QY 597 VDGIENTATTQLDVA-DIGSLSDLVSOVEVVISLLPASPHAAIAGVCIELKHHMTASVY 655
DB 520 LGKKYNINPVSMIDICKOEELGFLVAKQDLVISLLPVVHLPLVAKACITKNVNMVTSYI 579
QY 656 DESMSLSQAADAGVITLCEMGLDPCIDHLMKMKMIDEAHARKGKIKAFYSCGGLPSP 715
DB 580 TPALKLEKSVEDAGITILGELGLDPCGLHMLAMETIDKAKEVGATIESIYSCGGLPAP 639
QY 716 AAANPLAYKFSWNPAGALSGKNPAPVYKFLGETIIVDGHNLVYESAKRLRLRELPFALE 775
DB 640 EHSNNPLRYKFSWSPVGLVMNMQSATYLLDGKVVNVAGGISFLDAV-TSMDFPGLNLE 859
QY 776 HLPNRNSLIYGLYGISKEASTIYRATRYEGSEIMVTLKTGFFDAANHPLLOQTSRP 835
DB 699 GYPNRDSTKYAEIYGIS-SAHTLRLGTLRYKGYMKALNGFVKGLINREALPAFRPEANP 757
QY 836 -TYKGFDELNNISTINTDLIDIASGGYDDDLIARLLKLGCCCKNKEIAVTKTIFL 894
DB 758 LTKQLLCLDLVG-----ISPSSEHDLVKEAVLKLGDNQ-----LEAAEWLG 801
QY 895 LHEETQIPKGCSSPFDVICORMEORMAYGHNEQDMLLHHEVEVEYPDGQPAEKHQATLL 954
DB 860 AYGDI-NGFS--AMAKTVGLPTAMAAMLLDGEIGAKGLMGPFSEKIYGPILERIKARGI 916

RESULT 13
O95462 PRELIMINARY; PRT; 926 AA.
AC O95462;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE LYSINE-KETOGLUTARATE REDUCTASE /SACCHAROPINE DEHYDROGENASE.
GN LKR/SDH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Papes F., Kemper E.L., Cord-Neto G., Langone F., Arruda P.;
RT "Cloning and expression analysis of the LKR/SDH gene in human
tissues.";

Db 802 LLGDEQVQ-AESTLDALSKHLVWMLSYGPEEKDMIVMRDSFGIRHPSGH-LBHKITDLV 859
Qy 955 EFGKVENGRSTAMALTVGTPAATGALLLLKNKVKQTVIRPQEIYVPALFESSGI 1014
Db 860 AYGD1-NGFS--AWAKTVGLPTAAAKWMLDGEITGANGLMGPFSEKEYIGPILERIKAEI 916

RESULT 14
Q9C2G6 PRELIMINARY; PRT; 448 AA.
AC Q9C2G6;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PROBABLE SACCHAROPINE REDUCTASE (LVS3).
GN 93G11.270.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Algn V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL513443; CAC28679.1;
SQ SEQUENCE 448 AA; 49059 MW; 1EF38083F6CC1215 CRC64;

Query Match 14.1%; Score 746.5; DB 3; Length 448;
Best Local Similarity 35.8%; Pred. No. 4.3e-43;
Matches 173; Conservative 94; Mismatches 165; Indels 51; Gaps 11;

Qy 545 KILIGAGRCVCPAAEFLLASYPDICTYGVDDHDADQIHIVIVASLYQKDABETVPGIENTT 604
Db 6 KVLMLGAGFVTRPTLDVLE-----AGIPTVACRTLASAOKLSEGVKNAT 51
Qy 605 ATQLOVADIGSLDSLVQVEVVISLLPASFAAATAGVCIELKHKHVTASVDESMSLSQ 664
Db 52 PISLQVTDNDALDAEVAKHDLVISLIPYTHATVYKSAIROKKHVVTTSYVSPAMMELDA 111
Qy 665 AAKDAGVILCEMLDPCIDHLMMSKMTDEAHARKGKTKAFTSYCGGLSPAAANNPLAY 724
Db 112 EAKAAGITVMNEIGLDPGIDHLYAIAKTIDEVHQAGGKILSFSLCYGGLPAPEDSDNPLG 171
Qy 725 FSNWPAGALRSKGNPAAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNSLI 784
Db 172 FSNSSRGVLLALRNAASFYDKGKVTNAGPELMATAKPYFI--YPGAFVAVPNRDSSTI 229
Qy 785 YDLYGSKASTYRATXRYEGFSEIMVTLSTKTFDAAHPLLOQTSRPT-----YK 838
Db 230 YKERYNI--PEAQTVIRGTLRYQGPPQFIKTLVDIGFLD-----DTAESLSKQTPWK 280
Qy 839 GFLDELLNISTINTDIEASGCGYDDDLIARLLKCCCKNKEIAVTKVTKIFLGL-HE 897
Db 281 EATKEIVCAASSQADLE-----AAILSKATFESAEDQKRIILSGLRWIGLFS 328
Qy 898 ETQPKGSSPFVICORMEQRMAYGHNEQDMVLLHHEVEVEYDPGQPAEKHOATLLEFG 957
Db 329 ETITPRG--NPLDTLCATLEKMQFEGGERDLVMLQHKFEIHADGS--RETRTSTLVEYG 385
Qy 958 KVENGRSTTAMALTVGTPAATGALLLLKNKVKQTVIRPQEIYVPAL-EILESSGKIL 1016
Db 386 ---DPKGYSAWAKTVGVCAPAVKQVLSGQISGKVLAPMNSTDITEPLMKELHEKYGITM 442

Qy 1017 VEK 1019
Db 443 IEK 445

RESULT 15
Q9P4R4 PRELIMINARY; PRT; 450 AA.
AC Q9P4R4;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE SACCHAROPINE REDUCTASE.
GN LVS3.
OS Magnaporthe grisea (Rice blast fungus) (Pyricularia grisea).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
OX NCBI_TaxID=148305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4091-5-8;
RC MEDLINE=20235668; PubMed=10771443;
RA Johansson E., Steffens J.J., Emptage M., Lindqvist Y., Schneider G.;
RT "Cloning, expression, purification and crystallization of saccharopine
RL reductase from Magnaporthe grisea";
RL Acta Crystallogr. D Biol. Crystallogr. 56:662-664 (2000).
DR EMBL; AF144424; AAF91081.1;
SQ SEQUENCE 450 AA; 49059 MW; 699732A79C9E5C00 CRC64;

Query Match 13.6%; Score 717; DB 3; Length 450;
Best Local Similarity 35.6%; Pred. No. 4.6e-41;
Matches 170; Conservative 90; Mismatches 181; Indels 36; Gaps 10;

Qy 546 ILIIGAGRCVCPAAEFLLASYPDICTYGVDDHDADQIHIVIVASLYQKDABETVPGIENTT 605
Db 6 VLMGSGFVTRPTLDVLE-----SGIKVTACRTLESAAKLSAGVQHSSTP 51
Qy 606 TQLOVADIGSLDSLVQVEVVISLLPASFAAATAGVCIELKHKHVTASVDESMSLSQA 665
Db 52 ISLDVNDAAALDAEVAKHDLVISLIPYTHATVYKSAIROKKHVVTTSYVSPAMMELDA 111
Qy 666 AAKDAGVILCEMLDPCIDHLMMSKMTDEAHARKGKTKAFTSYCGGLSPAAANNPLAY 725
Db 112 AKDAGITVMNEIGLDPGIDHLYAIAKTIEGVAAGGKIKTFLSYCGGLPAPESDNPGLYK 171
Qy 726 FSNWPAGALRSKGNPAAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNSLI 785
Db 172 FSNSSRGVLLALRNAASFYDKGKVTNAGPELMATAKPYFI--YPGAFVAVPNRDSSTP 229
Qy 786 GDLYGSKASTYRATXRYEGFSEIMVTLSTKTFDAAHPLLOQTSRPTYKGFDEL 845
Db 230 KERYQI--PEADNIVRGLRYQGPPQFIKTLVDIGF-----LSDEQPP---FLKEAI 276
Qy 846 NNISTINTDIEASGCGYDDDLIARLLKCCCKNKEIAVTKVTKIFLGLHEETOI-PKG 904
Db 277 PKEA--TQIKVKAASASEQDIVSTIVSNATFESTEQKRIYAGLWGLGIFSKKITPRG 334
Qy 905 CSSPFVICORMEQRMAYGHNEQDMVLLHHEVEVEYDPGQPAEKHOATLLEFGKVENGRS 964
Db 335 --NALDTLCATLEKMQFEGGERDLVMLQHKFEIENKOGS--RETRTSLCEYGAIGSGG 391
Qy 965 TTAMALTVGTPAATGALLLLKNKVKQTVIRPQEIYVPAL-EILESSGKILVEK 1020
Db 392 YSAMAKFVGVPCAVAKFVLDGTISDRGLVAPMNSKINDPLMKELKEYIECKEV 448

Search completed: May 31, 2002, 15:51:48
Job time: 245 sec

us-09-049-304a-122.rspt

Tue Jun 4 15:23:17 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2002, 00:12:00 ; Search time 72.31 Seconds
(without alignments)
11091.044 Million cell updates/sec

Title: US-09-049-304A-120

Perfect score: 3265

Sequence: 1 ATGTGCGCGCCTTCTGCTA.....TCAATAATGCCGATCAGTT 3265

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105.2	3.2	323	1	US-08-474-633A-103
2	92.8	2.8	372	1	US-08-474-633A-102
3	45.4	1.4	7218	1	US-08-232-463-14
4	37.2	1.1	2659	1	US-08-007-775-3
5	37	1.1	3414	1	US-08-764-100-17
6	37	1.1	3414	1	US-08-764-100-21
7	37	1.1	4970	1	US-08-764-100-14
8	37	1.1	4970	1	US-08-764-100-20
9	36.4	1.1	5661	4	US-08-938-105-2
10	36.2	1.1	198	5	PCT-US95-10668-1
11	36.2	1.1	198	5	PCT-US95-10668-2
12	36.2	1.1	198	5	PCT-US95-10668-3
13	36.2	1.1	198	5	PCT-US95-10668-4
14	36	1.1	2105	4	US-09-088-425-3
15	36	1.1	2244	4	US-09-272-414-1
16	35.8	1.1	8010	4	US-09-521-526-2
17	35.8	1.1	8010	5	PCT-US95-11859-2
18	35.4	1.1	2700	3	US-09-022-875-1
19	34.8	1.1	2671	2	US-08-666-367B-1
20	34.8	1.1	2671	4	US-09-143-438-1
21	34.6	1.1	3489	2	US-08-728-323A-1
22	34.6	1.1	32207	2	US-08-770-379-20
23	34.6	1.1	32207	4	US-08-757-669A-20
24	34.6	1.1	32207	4	US-09-230-371A-20
25	34.4	1.1	558	2	US-08-732-791-2
26	34.4	1.1	558	3	US-09-205-049-2
27	34.4	1.1	1891	4	US-08-973-462-3

28	34.4	1.1	4376	1	US-08-119-125A-1
29	34.4	1.1	6744	1	US-08-119-125A-2
30	34.2	1.0	15328	2	US-08-888-497-33
31	34.2	1.0	15328	4	US-09-362-230-33
32	34.2	1.0	15328	5	PCT-US94-07926-33
33	34	1.0	3642	3	US-08-946-026-16
34	33.8	1.0	2080	1	US-08-179-738-9
35	33.8	1.0	2080	2	US-08-628-145-9
36	33.8	1.0	2257	1	US-08-171-718-15
37	33.8	1.0	2257	3	US-08-478-087-15
38	33.4	1.0	9785	4	US-09-479-128-1
39	33.2	1.0	1150	1	US-08-161-406-1
40	33.2	1.0	6038	4	US-09-305-639-4
41	33.2	1.0	7622	4	US-09-305-639-1
42	33.2	1.0	8532	1	US-08-452-655B-1
43	33.2	1.0	8532	3	US-08-450-582-1
44	33.2	1.0	8835	3	US-08-884-324-10
45	33.2	1.0	9606	1	US-07-741-940-1

ALIGNMENTS

RESULT 1
US-08-474-633A-103
; Sequence 103, Application US/08474633A
; Patent No. 5773691
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
; APPLICANT: COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND
; TITLE OF INVENTION: METHODS FOR INCREASING
; TITLE OF INVENTION: INCREASING THE LYSINE
; TITLE OF INVENTION: AND THREONINE CONTENT
; TITLE OF INVENTION: OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; ADDRESSEE: AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,633A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-474-633A-103

Query Match 3.2%; Score 105.2; DB 1; Length 323;
Best Local Similarity 65.1%; Pred. No. 5.3e-24;

	Matches	155;	Conservative	0;	Mismatches	83;	Indels	0;	Gaps	0;
QY	2844	ANGACCAAGCAGCAGCTACTGGAGTTTCGGGAGGTTGAAATGCGAGGTCCACCAATGCC	2903							
Db	3	AGCACACTGCCACTCTTTTGGAAATTCGGGACATCAAGATGCACAAACAACACCGCT	62							
QY	2904	ATGGCGCTCAGCCGTGGGCATTCAGCAGCAATAGGGGCCCTGCTATTTCCTAAAGATAAG	2963							
Db	63	ATGGCCAAAGACTGTTGGGATCCCTCAGCCCAATGGAGCTCTGCTTTAAATTGAAGACAAG	122							
QY	2964	GTCCACAGAAAGGAGTGATCAGGCTCTCGCAACCGGAAATCTACGTTCCAGCAATGGAG	3023							
Db	123	ATCAAGACAAAGAGGAGCTTTAAGGCTCTCGAAGCAGAGGTGATTTCGCCAGCTTTGGAT	182							
QY	3024	ATCTTGGAGTCGTGGGCATCAAGTGGTTGAGAAGTCGAGACTTGAAGCTTCCCTC	3081							
Db	183	ATATTGCAAGCATATCGTATAAGCTGTATGGAGAGGGCAGGAATGATCAAGAACATCG	240							

```

RESULT      2
US-08-474-633A-102
; Sequence 102, Application US/08474633A
; Patent No. 5773691
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
; COMPANY:
; TITLE OF INVENTION: CHIMERIC GENES AND
; TITLE OF INVENTION: METHODS FOR INCREASING
; TITLE OF INVENTION: INCREASING THE LYSINE
; TITLE OF INVENTION: AND THREONINE CONTENT
; TITLE OF INVENTION: OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; ADDRESSEE: AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,633A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIGHELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 372 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-08-474-633A-102

```

Query Match 2.8%; Score 92.8; DB 1; Length 372;
Best Local Similarity 61.1%; Pred. No. 6.3e-20;
Matches 179; Conservative 0; Mismatches 93; Indels 21; Gaps 1;
QY 1639 TTTTAAATCTTGGAGCTGGAAGATCTGTGGCCAGCTCTGTAGTGTCTGGCATCTTACC 1698

[illegible]

RESULT 3
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIELINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Davidson Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
; US-08-232-463-14

Query Match	1.4%	Score 45.4;	DB 1;	Length 7218;
Best Local Similarity	7.6%	Pred. No. 0.0016;		
Conservative	217;	Mismatches 198;	Indels	0; Gaps 0;

QY 2735 GATTTCGACGGAATGAACAGAGGATGCCTATGCCACAAATGAGCAAGACATGGTACT 2794
DB 1445 GAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1386
QY 2795 GCTCCACCAAGAGTCGAGGTGAATACCCGACGGCAACCGCCGAAAGACCAAGC 2854
DB 1385 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1326
QY 2855 GACGCTACTGGAGTTCGGGAAGTTGAAATGCGAGGTCCACCACTGCCATGGCGCTGAC 2914
DB 1325 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1266
QY 2915 CTTCCGCGATTCAGCAGCAATAGGGGCCCTCTATTGCTAAAGATAAGTCCAGACGAA 2974
DB 1265 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1206
QY 2975 AGGAGTATGAGGCTCTGCAACCGAAATCTAGTTCAGCATTTGGAGATC 3034
DB 1205 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1146
QY 3035 GTCGGGCATCAAGCTGGTGAAGAGTGAGACTTGAAAGTTCCCTGATACAGATAA 3094
DB 1145 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1086
QY 3095 CATAGTATGATAGCAGGCGACATGTATCTTTGTATTAACTCCGTTCTGGAATATA 3154
DB 1085 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1026
QY 3155 TTGTGAACTAAATGTGACAAATAAAA 3183
DB 1025 TCTGTGCGGTATGGCAACGAAGGAAA 997

RESULT 4

US-08-007-775-3
; Sequence 3, Application US/08007775
; Patent No. 5340733
; GENERAL INFORMATION:
; APPLICANT: Takashi UENO et al.
; TITLE OF INVENTION: MBOI RESTRICTION-MODIFICATION GENES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/007,775
; FILING DATE: 19930122
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2659 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double

; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Moraxella bovis
; STRAIN: 10900
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL LINE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION: 12
; IDENTIFICATION METHOD:
; OTHER INFORMATION: /note="Inosine"
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
US-08-007-775-3

Query Match 1.1%; Score 37.2; DB 1; Length 2659;
Best Local Similarity 51.4%; Pred. No. 0.34;
Matches 110; Conservative 0; Mismatches 103; Indels 1; Gaps 1;
QY 1679 TGAGTTTCTGGCATCTTACCCAGACATATGTACCTATGTTGTTGATGACCATGATGCAGA 1738
DB 766 TGAGAGAATAACAAGCTAACCGCACCATTTAGTCCCAATAGTAATGTCGCAAAAAGGTAAA 825
QY 1739 TCAAAATTCATGTTATCTGTTGGCATCTTTGTATCAAA-AGATGCAGAAAGACACAGTTGATG 1797
DB 826 TCAAAATTCATGTTATCTGTTGGCATCTTTGTATCAAA-AGATGCAGAAAGACACAGTTGATG 1797
QY 1798 GTATTGAAAATACAACTGCTTACCCAGCTTTGATGTTGCTGATATGGAAGCTTTTCAGATC 1857
DB 886 GTATGTCAGAAACCAATACTACTTTGGATTAATTTTACTGTTTGTATGATAAAGTAAAAAAGA 945
QY 1858 TTGTTTCTCAGGTTGAGTTGTTGTAATTTAGCTTGCT 1891
DB 946 ATGTCGCTCAAAATTGAGATTCACTTAAACCCAGCT 979

RESULT 5
US-08-764-100-17/c
; Sequence 17, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gielen L., Johannes J.
; APPLICANT: Peters, Dirk
; APPLICANT: Goldbach, Robert W.

ADDRESSEE: Sandoz Agro, Inc
STREET: 975 California Avenue
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,100
FILING DATE: 06-DEC-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,064
FILING DATE:

APPLICATION NUMBER: US 08/032,235
FILING DATE: 17-MAR-1993
APPLICATION NUMBER: GB 9206016.9
FILING DATE: 19-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: NO. 57737001s, Allen E.
REGISTRATION NUMBER: 34,490
REFERENCE/DOCKET NUMBER: 137-1061
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 354-3592
TELEFAX: (415) 857-1125

INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 3414 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-764-100-21

Query Match 1.1%; Score 37; DB 1; Length 3414;
Best Local Similarity 53.0%; Pred. No. 0.49;
Matches 79; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

[illegible]

RESULT 7
 US-08-764-100-14/c
 ; Sequence 14, Application US/08764100
 ; Patent No. 5773700
 ; GENERAL INFORMATION:
 ; APPLICANT: van Grinsven J., Martinus Q.
 ; APPLICANT: De Haan, Petrus T.
 ; APPLICANT: Gielen L., Johannes J.
 ; APPLICANT: Peters, Dirk
 ; APPLICANT: Goldbach, Robert W.
 ; TITLE OF INVENTION: Improvements in or Relating to Organic
 ; TITLE OF INVENTION: Compounds
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sandoz Agro, Inc
 ; STREET: 975 California Avenue
 ; CITY: Palo Alto
 ; STATE: CA

RESULT 6
 US-08-764-100-21
 ; Sequence 21, Application US/08764100
 ; Patent No. 5773700
 ; GENERAL INFORMATION:
 ; APPLICANT: van Grinsven J., Martinus Q.
 ; APPLICANT: De Haan, Petrus T.
 ; APPLICANT: Gielen L., Johannes J.
 ; APPLICANT: Peters, Dirk
 ; APPLICANT: Goldbach, Robert W.
 ; TITLE OF INVENTION: Improvements in or Relating to Organic
 ; TITLE OF INVENTION: Compounds
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:

RESULT 7
US-08-764-100-14/c
; Sequence 14, Application US/08764100
: Patent No. 5773700

COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,100
FILING DATE: 06-DEC-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,064
FILING DATE: 17-MAR-1993
APPLICATION NUMBER: GB 9206016.9
FILING DATE: 19-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 5773700r1s, Allen E.
REGISTRATION NUMBER: 34,490
REFERENCE/DOCKET NUMBER: 137-1061
TELEPHONE: (415) 354-3592
TELEFAX: (415) 857-1125
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 4970 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-764-100-14

Query Match 1.1%; Score 37; DB 1; Length 4970;
Best Local Similarity 53.0%; Pred. No. 0.65;
Matches 79; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 1694 TTACCAGACATATGCTATGCTGTGATGACCATGATGACATCAAAATTCATGTTAT 1753
Db 4756 TCATCCAGAAATTTATGATGAAGCTTATTGACCGTCTGTAGATCATAAATGAAT 4697
Qy 1754 CGTGGCATCTTTGATCAAAAGATGCAGAGACAGTGTGATGTTGAAATACAAC 1813
Db 4696 TCTAGATACCTTTGGCTGAAATGCTCCAGAATGCAACAGGTAACACCTTAGACCAACAG 4637
Qy 1814 TGCTACCCAGCTTGATGCTGTGCTGATATTG 1842
Db 4636 AGATACTCAAACTCTTCTGCCATAATG 4608

RESULT 8
US-08-764-100-20
Sequence 20, Application US/08/764100
Patent No. 5773700
GENERAL INFORMATION:
APPLICANT: van Grinsven J., Martinus Q.
APPLICANT: De Haan, Petrus T.
APPLICANT: Gielen L., Johannes J.
APPLICANT: Peters, Dirk
APPLICANT: Goldbach, Robert W.
TITLE OF INVENTION: Improvements in or relating to Organic
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sandoz Agro, Inc
STREET: 975 California Avenue
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COUNTRY: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,100
FILING DATE: 06-DEC-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,064
FILING DATE: 17-MAR-1993
APPLICATION NUMBER: GB 9206016.9
FILING DATE: 19-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 5773700r1s, Allen E.
REGISTRATION NUMBER: 34,490
REFERENCE/DOCKET NUMBER: 137-1061
TELEPHONE: (415) 354-3592
TELEFAX: (415) 857-1125
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 4970 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-764-100-20

Query Match 1.1%; Score 37; DB 1; Length 4970;
Best Local Similarity 53.0%; Pred. No. 0.65;
Matches 79; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 1694 TTACCAGACATATGCTATGCTGTGATGACCATGATGACATCAAAATTCATGTTAT 1753
Db 215 TCATCCAGAAATTTATGATGAAGCTTATTGACCGTCTGTAGATCATAAATGAAT 274
Qy 1754 CGTGGCATCTTTGATCAAAAGATGCAGAGACAGTGTGATGTTGAAATACAAC 1813
Db 275 TCTAGATACCTTTGGCTGAAATGCTCCAGAATGCAACAGGTAACACCTTAGACCAACAG 334
Qy 1814 TGCTACCCAGCTTGATGCTGTGCTGATATTG 1842
Db 335 AGATACTCAAACTCTTCTGCCATAATG 363

RESULT 9
US-08-938-105-2
Sequence 2, Application US/08938105
Patent No. 6353151
GENERAL INFORMATION:
APPLICANT: Leinwand, Leslie A.
APPLICANT: Vikstrom, Karen L.
TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,105
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:

NAME: Crook, Wannell M.
REGISTRATION NUMBER: 31,071
REFERENCE/DOCKET NUMBER: 3595-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5661 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..5661
US-08-938-105-2

Query Match 1.1%; Score 36.4; DB 4; Length 5661;
Best Local Similarity 51.9%; Pred. No. 1.1; Indels 0; Gaps 0;
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 2720 GAGCCCATTTGATGTTGCCAGCAATGGAACAGAGGATGGCCTATGGCCACAATGA 2779
DB 4356 GRACTCCAGGAGGAGATCTCAGACTGACTGAACAGCTGGGAGAGGGGTAAAAATGT 4415
QY 2780 GCAAGACATGTTACTCTCCACCACGAAGTCGAGGTGGATACCGGACGCGCAACCCGC 2839
DB 4416 GCACGAGCTGGAGAGATCCGAAACAGCTGGAGGTGGAGAGCTGGAACTGCAGTCAGC 4475
QY 2840 CGAAAAGCACCAGCGACGCTACTGTGAGTTCGGGAAGG 2877
DB 4476 CCTGGAGGAGCTGAGGCCCTCCCTGAGCATGAGGAGG 4513

RESULT 10
PCT-US95-10668-1/c
Sequence 1. Application PC/TUS9510668
GENERAL INFORMATION:
APPLICANT: James Eberwine
TITLE OF INVENTION: A Method of Sequencing Proteins by
TITLE OF INVENTION: Epitope Ordering and Protein
TITLE OF INVENTION: Restriction Mapping
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jane Massey Licata, Esq.
STREET: 210 Lake Drive East, Suite 201
CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
ZIP: 08002
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294,133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 198
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: NO
PCT-US95-10668-1

Query Match 1.1%; Score 36.2; DB 5; Length 198;
Best Local Similarity 50.9%; Pred. No. 0.096;
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 1715 TGGTGTGATGACCATGATCCAGATCAAAATTCATGTTATCGTGGCATCTTGTATCAAAA 1774
DB 170 TGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 111
QY 1775 AGATGCAGAGAGACAGCTTCATGTTATGAAATACAACTGCTACCCAGCTTGATGTTGC 1834
DB 110 TGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 51
QY 1835 TGATATTGGAAGCTTTCAGATCTTGTTCAGTTCTCAGTTGAGTTGATTAATT 1883
DB 50 TGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2

RESULT 11
PCT-US95-10668-2/c
Sequence 2. Application PC/TUS9510668
GENERAL INFORMATION:
APPLICANT: James Eberwine
TITLE OF INVENTION: A Method of Sequencing Proteins by
TITLE OF INVENTION: Epitope Ordering and Protein
TITLE OF INVENTION: Restriction Mapping
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jane Massey Licata, Esq.
STREET: 210 Lake Drive East, Suite 201
CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
ZIP: 08002
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294,133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 198
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: NO
PCT-US95-10668-2

Query Match 1.1%; Score 36.2; DB 5; Length 198;
Best Local Similarity 50.9%; Pred. No. 0.096;
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

[illegible]

RESULT 12

```

PC97-US95-10668-3/c
; Sequence 3, Application PC/TUS9510668
; GENERAL INFORMATION:
; APPLICANT: James Eberwine
; TITLE OF INVENTION: A Method of Sequencing Proteins by
; TITLE OF INVENTION: Epitope Ordering and Protein
; TITLE OF INVENTION: Restriction Mapping
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jane Massey Licata, Esq.
; STREET: 210 Lake Drive East, Suite 201
; CITY: Cherry Hill
; STATE: NJ
; COUNTRY: USA
; ZIP: 08002

```

```

? KIP: 0800Z
? COMPUTER READABLE FORM:
? MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
? COMPUTER: IBM 486
? OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
? SOFTWARE: WORDPERFECT 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US95/10668
? FILING DATE: Herewith
? CLASSIFICATION:

```

PRIORITY: 3
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/294,133
 FILING DATE: August 22, 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Jane Massey Licata
 REGISTRATION NUMBER: 32,257
 REFERENCE/DOCKET NUMBER: PENN-0137
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (609) 779-2400
 TELEFAX: (609) 779-8488
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:

; LENGTH: 198

```

TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: NO
PCT-US95-10668-3

```

```

Query Match      1.1%; Score 36.2; DB 5; Length 198;
Best Local Similarity 50.9%; Pred. No. 0.096;
Matches 86; Conservative 0; Mismatches 83; Indels

```

[illegible]

Db 50 TGTTGTTGGAATCTTGTGTTGTTGTTGTTGTTGTTGTT 2

RESULT 13

```

PCR-US95-10668-4/c
; Sequence 4, Application PC/TUS9510668
; GENERAL INFORMATION:
; APPLICANT: James Eberwine
; TITLE OF INVENTION: A Method of Sequencing Proteins by
; TITLE OF INVENTION: Epitope Ordering and Protein
; TITLE OF INVENTION: Restriction Mapping
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Jane Massey Licata, Esq. 201
; STREET: 210 Lake Drive East, Suite 201
; CITY: Cherry Hill
; STATE: NJ
; COUNTRY: USA
; ZIP: 08002

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294,133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 198
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; ANTI-SENSE: NO
PCT-US95-10668-4

```

Query Match 1.1%; Score 36.2; DB 5; Length 198;
Best Local Similarity 50.9%; Pred. No. 0.096;
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

1715	TGGTGTGATGACCATGATGAGATCAAATTCATGTTATCGTGCATCTTTGTATCAAAA	1774
QY		
170	TGTTGTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	111
DB		
1775	AGTCGAGAAGACAGCTGTGATGTGATTTGAAATACAACTGCTACCCAGCTTGATGTTGC	1834
QY		
110	TGTTGTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	51
DB		

1835'	TGATATTGGAGCCCTTTCAGATCTTGTTCACGTTGAGGTGTGAATT	1883
20Y		
50	TGTTGTTGGAAATTCTGCTTGTGTGTTGTTGTTGTTGTTGTTGTT	2
dbb		

RESULT 14

Sequence 3, Application US/09088425
Patent No. 6171843
GENERAL INFORMATION:
APPLICANT: BANDMAN, OLGA
APPLICANT: LAL, PRETI
APPLICANT: CORLEY, NEIL C.

APPLICANT: PATTERSON, CHANDRA
APPLICANT: BAUGHN, MARIAH R.
TITLE OF INVENTION: HUMAN ISOMERASE-LIKE PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/088,425
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Cerrone, Michael C
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0529 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2105 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAINON01
CLONE: 2291164
PS-09-088-425-3

	Query Match	1.1%	Score 36;	DB 4;	Length 2105;	
	Best Local Similarity	53.6%	Pred. No.	0.7;		
Matches	75; Conservative	0;	Mismatches 65;	Indels 0;	Gaps 0:	
QY	1703 CATATGTACCTATGGTGTGATGACCATGATCAGATCAAAATTGTTATCGTGCGCATC	1762				
Ddb	890 CAAAAGTAGTCATGACTGCCTTAAGNATGATCACATCTCAGTCTGTTCCAGTTGTAGA	949				
QY	1763 TTTGTATCAAAAAGATCAGAAGAAGACAGTTGATGGTATTGAANAATACAACGTGCTACCCA	1822				
Ddb	950 AAGTCAAAAAGGTGATGCCACCAATTATTGTTGATGGAGAAGATCAAAGTGCAGAGCA	1009				
QY	1823 GCTTGAATTTGCTGATATTG	1842				
Ddb	1010 TGATGAATAATATTGATGGTG	1029				

```

RESULT 15
US-09-272-414-1
; Sequence 1, Application US/09272414A
; Patent No. 6238885
;
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola G.
; TITLE OF INVENTION: Histidine Kinase
; FILE REFERENCE: GM10202
; CURRENT APPLICATION NUMBER: US/09/272,414A
; CURRENT FILING DATE: 1999-03-19
;
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1

```

```

; LENGTH: 2244
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-272-414-1

Query Match      1.1%; Score 36; DB 4; Length 2244;
Best Local Similarity 51.2%; Pred. No. 0.74;
Matches 84; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1464 GAAGTAGGAGCAGATGATACGTCCACATTTGGATTAATATTGATTCCTTGACATCTTTA 1523
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 gaagaagcaataaataatagtcacaacttaattgaatacctgtgggttgatgattata 338

QY 1524 GCTAATGNACATGCTGGAGATCAGATGCCGGCAAGAAATTGAATTAGCTCTCAAGATA 1583
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 aataaataacatcgtcaatcaacggtctcaattctataataaagaacaaatttgaaat 398

QY 1584 GGAAAGTCAATAGTATGAAACTGACGCTACAAATTGTAAGAAG 1627
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 199 gaadgaactcaacagcaccattttgaacatgtgtttgataaagg 442
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: June 1, 2002, 02:05:08
Job time: 6788 sec



GenCore version 4.5.
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2002, 00:15:00 ; Search time 346.29 Seconds
(without alignments)
16187.951 Million cell updates/sec

Title: US-09-049-304a-120

Perfect score: 3265

Sequence: 1 APTGTGCCCGCTTCGTCTA.....TCAATATGCGGATCAGT 3265

Scoring table:

IDENTITY-NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3263	99.9	3265	20	Maize lysine ketog
2	1908	58.4	1908	20	Maize lysine ketog
3	1233.2	37.8	3195	21	Arabidopsis thalia
4	907	27.8	2582	20	Soybean lysine ket
5	522.6	16.0	720	20	Rice lysine ketog
6	515	15.8	1449	20	Arabidopsis saccha
7	183.6	5.6	429	20	Wheat lysine ketog
8	155.8	4.8	2919	23	Drosophila melanog
9	155.8	4.8	3266	22	Drosophila melanog

C	10	110.4	3.4	6209	23	ABL12396
	11	105.2	3.2	323	19	AAV35857
	12	105.2	3.2	323	20	AAV95554
	13	101	3.1	308	20	AAV95566
	14	97.6	3.0	1262	21	AAF13730
	15	95.8	2.9	8160	20	AAV95559
	16	92.8	2.8	372	19	AAV35856
	17	92.8	2.8	372	20	AAV95553
	18	79.4	2.4	598	21	AAF07883
	19	49.2	1.5	4590	22	AAH24065
	20	41.6	1.3	1830	23	AAS33006
C	21	41.6	1.3	2717	20	AAI13034
	22	41.4	1.3	2331	22	AAH16467
C	23	41.4	1.3	4840	23	ABL16917
	24	41.4	1.3	18189	23	ABL16916
	25	41.2	1.3	556	22	ABA62873
	26	41.2	1.3	556	22	ABA30153
	27	41.2	1.3	556	22	AAI17913
	28	41.2	1.3	556	22	AAI42893
C	29	40.4	1.2	2693	19	AAV71119
	30	40	1.2	608	24	AB199273
C	31	40	1.2	4590	22	AAH24065
	32	39.8	1.2	2157	23	AA569144
C	33	39.6	1.2	486	22	AAF75507
	34	39.6	1.2	2169	21	AAA30203
C	35	39.6	1.2	15016	20	AAK99560
	36	39.2	1.2	531	22	AAK32668
C	37	39.2	1.2	531	22	AAK34883
	38	38.8	1.2	6558	21	AAZ45602
C	39	38.6	1.2	3114	23	ABL26943
	40	38.6	1.2	5114	23	ABL26942
	41	38.6	1.2	11046	24	AA516913
	42	38.4	1.2	222	22	ABA70505
	43	38.4	1.2	222	22	ABA37133
	44	38.4	1.2	222	22	AAK18749
	45	38.4	1.2	222	22	AAK44685

ALIGNMENTS

RESULT 1

AAV99563

ID AAV99563 standard; cDNA to mRNA; 3265 BP.

XX AAV99563;

AC AAV99563;

XX 29-MAR-1999 (first entry)

XX Maize lysine ketoglutarate reductase cDNA.

DE Maize lysine ketoglutarate reductase cDNA.

XX Lysine ketoglutarate reductase; saccharopine dehydrogenase;

KW transgenic plant; seed; soybean; ss.

XX Zea mays.

OS Zea mays.

XX Key

XX Location/Qualifiers

FT CDS

3..3071

/*tag= a

XX WO9842831-A2.

XX 01-OCT-1998.

XX 27-MAR-1998;

XX 98WO-US06051.

XX 27-MAR-1997;

XX 97US-0824627.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Epelbaum SU, Falco SC, McDevitt RE;

XX WPI; 1999-045139/04.

Drosophila melanog
cDNA encoding prot
Arabidopsis lysine
Rice lysine ketogl
Aspergillus oryzae
Arabidopsis lysine
cDNA encoding prot
Arabidopsis lysine
Fusarium venenatum
Yeast AOD9604-asso
Enterococcus faeca
Enterococcus faeca
Human cDNA sequenc
Drosophila melanog
Drosophila melanog
Human foetal liver
Probe #8619 for ge
Probe #7846 for ge
Probe #11579 used
Exemplary caffeic
Mouse ischaemic co
Yeast AOD9604-asso
DNA encoding novel
Polyglutamine trac
Human RING finger
Nucleic acid sequ
Human bone marrow
Probe #7169 used t
cDNA sequence of a
Drosophila melanog
Drosophila melanog
Human G-protein co
Human foetal liver
Probe #15599 for g
Human brain expres
Human bone marrow

DR P-PSDB; AAW87761.

XX Nucleic acids and chimeric genes for increasing seed lysine content
PT - comprise sequence encoding all or part of lysine ketoglutarate
PT reductase, useful to improve nutritional quality of seeds from
PT transformed plants

XX Claim 3; Page 184-186; 231pp; English.

XX This is the nucleotide sequence of a near full-length cDNA encoding
CC maize lysine ketoglutarate reductase (LKR, see AAW87761), or
CC saccharopine dehydrogenase (SDH). It was obtained by PCR
CC amplification of developing seed cDNA using primers (see AAW99560-61)
CC based on Arabidopsis LKR/SDH sequences, and by RACE. Isolated
CC nucleic acids comprising sequences encoding all or part of plant
CC LKRs are new. Also claimed are: (1) a chimeric gene comprising the
CC fragment (or a subfragment) operably linked to a seed specific
CC regulator, where the chimeric gene reduces LKR activity in plant
CC seeds transformed with it; (2) plant cells and seeds in which LKR
CC activity is reduced due to a mutation in the gene encoding LKR or
CC transformation with the chimeric gene; (3) a nucleic acid fragment
CC comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC synthase (DHPS), substantially insensitive to lysine inhibition is
CC operably linked to a plant chloroplast transit sequence and to a
CC plant seed-specific regulatory sequence; (4) plants comprising in
CC genome (1) and (ii), especially as fragment of (2); and (5) seeds
CC from (3). The chimeric genes can be used to produce plant cells
CC and seeds with reduced LKR activity, especially in Arabidopsis,
CC corn, soybean, rapeseed, wheat and rice (claimed). LKR is
CC important in controlling free lysine accumulation in plant seeds.
CC LKR activity reduction may be achieved by cloning the claimed
CC fragment, preparing a chimeric gene for cosuppression of LKR,
CC expression of antisense RNA for LKR, and transforming plants with
CC the chimeric gene.

XX Sequence 3265 BP; 952 A; 651 C; 764 G; 897 T; 1 other;

Query Match 99.9%; Score 3263; DB 20; Length 3265;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTTGGCCCGCTTCTGCTAGGAGGAGGCAAGAACGCGACCTCAGTAAACCGGATTTTG 60
DB 1 attgtgcccgccttctgtagagaggaggaagaaacgagcctcagtaaacgagattatg 60
QY 61 TGCAGCCCAAGCAACAGGAGGATCCATCATGACGCTCAGTATGAGGATGAGGATGCGAGA 120
DB 61 tgcagccaagcaacagggagggatccatcatgacgctcagtagagagatgcaggtgcaga 120
QY 121 TTTCAGAGACCTGTGAGAAATGCGGCTTATCATAGGCATCAACAAACCCAAAGCTGCAGA 180
DB 121 ttfcagagacctgtcagaatgcgagcttcttccacacacacaaagcctgcaga 180
QY 181 TGATTTCTTTCAGATAGCGGTACGCTTTCTTTCACACACACAAAGCCCAAAAGAGA 240
DB 181 tgattcttcagatagagcgtacgcttcttccacacacacaaagcctgcaga 240
QY 241 ATATGCCACTGTAGACAGATCCCTTGAAGAAAGGCTCCTTGTGATTTATGAGCTAA 300
DB 241 atatgccactgttagacaagatcccttgaagaagggtctccttgttattatgagctaa 300
QY 301 TTGTTGAGATGATGGGAAAGATCAGTACGATTTGGGAAATTTGCTGGTAGCTGGAC 360
DB 301 ttgttggagatgatgggaaagatcactagcatttgggaaatttgcctggtgagctggac 360
QY 361 TGATAGATTTCTTACATGCTCTCGGACAGCATATTTAGCCCTTGGATCTCGACTCCAT 420
DB 361 tgatagattcttcatatgctctcggacagcagatattttagccttggactactgcacctat 420
QY 421 TTCTCTCTCTGGGACAATCTCATATGATCTTCCTCGCTCCCTGAGCCAGGCTGCAGTCA 480
DB 421 ttctctctctctgggacaatctcatatgattcttccctcgctccctgagccaggtgcagtc 480

DB 421 ttctctctctctgggacaatctcatatgtatctctctctctctctctctctctctctctctca 480
QY 481 TTGTCGTTGCGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCGATAGTGT 540
DB 481 ttgtcgttgcgaagagagatagcaacatttgactccatccggaatttgcgagatagt 540
QY 541 TTGTCGTTGCGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCGATAGTGT 600
DB 541 ttgttctcactggagtggaaacgtctctcaggtggtgcgagagatattcaagttatgtgc 600
QY 601 CCCATACCTTTGTTTTCATGCTGAGAGCTTCCGAAATTTTTCAGGCGCAGGAATCTGTCTA 660
DB 601 cccataacctttgttgatgctgagaagcttccgaaatttttcaggccaggaatctgtcta 660
QY 661 AGCAATCTCAGTCGACCAAGAGAGATTTTCAACTTTTATGTTGTTGTTGTTGTTGTTGTTG 720
DB 661 agcaatctcagtcgacccaagagattttcaactttatggtttgtgtgacctctagag 720
QY 721 ACATGTTTCTCACAAGGATCCACAGACATTTTGAAGAAATTTGCTCCATATGATGCTATC 780
DB 721 acatgtttctcacaagagatccacagacaatttgcacaaaggtgactattatgtctatc 780
QY 781 CAGAACACTACACCCCTGTTTTCATGAAAGAAATTTGCTCCATATGATGCTATCCTGTA 840
DB 781 cagaacactacacccctgttttctcagaagaattgtccatcatgcatctgcatctgtaa 840
QY 841 ACTGTATGTTGGGAGAGAGGTTTCCACCATTTACTAAATATGATGATGATGATGATGAT 900
DB 841 actgtatgtattggagaagaggtttccaccatctacaaatggatcagttacagcaat 900
QY 901 TGATGGAGACTGTTGCTTTAGTCGGGCTTTGTGACATACTTGTGATATGAGGTT 960
DB 901 tgatggagactggtgtcttttagtcgggctttgtgacatacttgtgatatgtgaggt 960
QY 961 CCATTGAAATTTTCAACAAGAGTACATCATAGAGAGGCTTTCTTTCGCTGATGATCCT 1020
DB 961 ccattgaaatttatacaacaagagatcataatagagagggccttcttcttcgtagatcctt 1020
QY 1021 CTAAGAATTCATACCATGATGATATGAAAGTCCCGAGTGCTGCTTGGCTGTTGACA 1080
DB 1021 ctaagaattcataccatgatatgataggagtgccgagtggtctgttctgttgcgttgaca 1080
QY 1081 TTCCTCCCTACAGAAATTTCTTAAGAGGCTCCCAACATTTTGGAAACATACTACTAGAC 1140
DB 1081 ttctccctacagaattctctaaagagcctcccaacatttggaaacatactactatagac 1140
QY 1141 TTGTTGCTAGTTTGGCTCAGTGAAGCAACCGCAGAACCTTCCTTCTACTTGGAGAGAG 1200
DB 1141 ttgttgcctagtttggcctcagtgaaagcaaccgagaaacttcttcttacttgaagaag 1200
QY 1201 CTTGCAATTCACATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGA 1260
DB 1201 cttgcattgcacatgctggcagattaaactcctttgtatgaataatacctctaggaagaa 1260
QY 1261 ATACTATGATAGATTGGCACCCGCAAAACAAATTCCTTCCTGCAAGAAAGTATAGCA 1320
DB 1261 atactatgatagatttggcacccgcgcaaaacaaatcccttgcctgacaagaagtataga 1320
QY 1321 CCGTGTATCTCTCAGTGGGACACCTATTGTAAGTTCCTTTATTAATCAAGCTTTGGACA 1380
DB 1321 cctgtgtatctctcagtggtggccacctatttgataaagttcctataaataagagctttggaca 1380
QY 1381 TCATTGAGACAGCTGGAGGTTTCATTTCACTTGTGTTAGATGTGAAGTTGGACAAAGCAGG 1440
DB 1381 tcattgagacagctggaggttcttcttcaacttggttagatgtgagttggacaaagcag 1440
QY 1441 ATGATATGTCATCTACTCAGAGCTTGAAGTAGGAGCAGATCATCTGCCACATTTGGATAAA 1500
DB 1441 atgatatgtcatactcagagcttgaagtaggagcagatgatactgccacattggataaaa 1500
QY 1501 TTATTGATTCCTTCACTTCTTTAGCTTAATCAACATGGTGGAGATACGATCCGGCGAAG 1560
DB 1501 ttattgattctctgactcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1560

QY 1561 AAATTGAATTAGCTCTGAAGATAGCAAAAGTCAATGATGATGAACCTGACGTGACAAATTC 1620
DB 1561 aaattgaattagctctgaagatagcaaaagtcaatgagtagtaaaactgacgtgacaaattg 1620
QY 1621 ATAAAGAGGCCAAAGATTTTAATCTTGAGCTGGAAGAGTCTGTCGGCCAGCTGCTG 1680
DB 1621 ataaagagggccaaagattttaattcttgagctggaagagctctgctggccagctgctg 1680
QY 1681 AGTTTCTGGCATCTTACCCAGACATATGTACCTATGCTGTTGATGACCATGATGATGATC 1740
DB 1681 agtttctggcatcttaccagacatattgtacctatggtttgatgaccatgatgcagatc 1740
QY 1741 AAATTCATGTTATGCTGGCATCTTTGATCAAAAAGATGAGAGAGACAGATTTGATGGTA 1800
DB 1741 aaatctcatgttatcggtggcatctttgtatcaaaaagatgagagagacagttgagtga 1800
QY 1801 TTGAAAATACAACTGTACCCAGCTTGATGTTGCTGATATGGAAGCCCTTTGAGATCTTG 1860
DB 1801 ttgaaaatacaactgtaccagcttgatgttgctgatatgttgaaagcctttcagatcttg 1860
QY 1861 TTCTCAGGTTGAGGTTGTAATAGCTTGCCTGCTAGTTTTCATGCTGCCATTTGCAG 1920
DB 1861 ttctcaggttgaggttgtaattagcttgcctgctgctagtttctatgctgcatgtgcag 1920
QY 1921 GAGTATGCATAGATTGAAGAAGCACATGTTAACGGCAAGCTATGTTGATGAATCCATGT 1980
DB 1921 gagtatgcatagatttgaagaagcacatggttaacggcaagctatgttgatgaatccatgt 1980
QY 1981 CAAACTTGAGCCAAAGCTGCCAAGATGCGAGCTGTAACTATCTTTGTGAAATGGCCCTAG 2040
DB 1981 caaacttgagccaaagctgccaagatgcaggtgttaactatactttgtgaaatggccctag 2040
QY 2041 ATCTCGGCATAGATCACTTGTATGTCATGAAGATGATTGATCAAGCTATGCAACCAAGG 2100
DB 2041 atctctgcatagatcaacttgatgtcaatgaagatgattgagaagctcatgcacgaaagg 2100
QY 2101 GAAAATAAAGGCATTTACATCTTACTGTGGTGGATGCCATCTCCAGCTGCGAGCAACA 2160
DB 2101 gaaaataaaggcatttacatcttactgtgggtggattgccatctccagctgcagcaaca 2160
QY 2161 ATCCGCTTGCCCTATAAATTCAGTTGGAACCCAGCTGGTGCCTCCGCTCAGGGAAATC 2220
DB 2161 atccgcttgccctataaattcagttggaacccagctgggtgcctccgctcaggggaaatc 2220
QY 2221 CTGCAGCTACAAATTTCTGTGAGACGATCCATGTTAGATGGTCAATACTGTATGAAT 2280
DB 2221 ctgcagctacaaatcttctgtgagacgatccatgttagatgggtcaataactgtatgaat 2280
QY 2281 CAGCAAAGAGGCTCAGACTACGAGACTTCCAGCTTTTGGCTCTGGAACACTTGGCAATC 2340
DB 2281 cagcaaagaggctcagactacgagacttccagcttttggctctggaacacttggcaaatc 2340
QY 2341 GGAATTCCTGATATATGGTGACCTTTATGTTATCTCCAAAGAAGCATCCACCATATATA 2400
DB 2341 ggaattcctgatatatgttgacctttatggtatctccaaagaagcatccaccatata 2400
QY 2401 GGGCTACTTTCCTTACCAAGCTTTAGTGAGATTATGGTAACCTTTCCAAAACTGGGT 2460
DB 2401 gggctacttctcttacgaaggttttagtgagattatggtaaacctttccaaaactgggt 2460
QY 2461 TCTTTGATGTCGAATCATCTGCTGCAAGTACTAGTCGTCGCAACATATAAGGGTT 2520
DB 2461 tctttgatgctgcaaatcatccactgctgaagatactagtcgtcccaacataaagggtt 2520
QY 2521 TCCCTTGATGAATCTACTGAATAATATCTCCAAATTAACACGGACTTATGATTTGAGCTT 2580
DB 2521 tcccttgatgaatctactgaataatctccaaatlaaacacggactttagatgaagctt 2580
QY 2581 CTGCTGGATGATGATGACTGATGTCAGACTGTTGAAGCTCGGGTGTGCAAAAATA 2640
DB 2581 ctgctggatgatgactgactgactgctgagactggtgaagctcggtgttgcaaaaata 2640

QY 2641 AGGAAATAGCTCTTAAAGACAGTCAAACCATCAAGTTCTTGGACTACATGATGAAGAGACTC 2700
DB 2641 aggaaatagctctttaaagacagtcaaaccatcaagttcttggactacatgatgaagagactc 2700
QY 2701 AAATACCTTAAGGTTTGTTCGAGCCCATTTGATGTGATTTGCCAGGCAATGGAACAGAGTA 2760
DB 2701 aaataccttaagggttgttcgagcccatcttgatgtgatttgcagcgaaatggaaacagagga 2760
QY 2761 TGGCCTATGGCCACAAATGAGCAAGACATGGTACTCTCCACACGAATCGAGGTGGAAT 2820
DB 2761 tggcctatggccacaaatgagcaagacatggttactgctccaccagaaagtcgaggtggaaat 2820
QY 2821 ACCCGAGCGGCAACCCGCCAAAGACCAAGCGACGCTACTTGGAGTTCCGGGAAGGTTG 2880
DB 2821 acccgagcgggcaacccgccaaagacccaagagcaactactggagttcggggaagggtg 2880
QY 2881 AAAATGGCAGGTCCACCACTGCCATGGCGCTGACCGTCCGCAATCCAGCAGCAATAGGG 2940
DB 2881 aaaatggcaggtccaccactgccatggcgctgacctggcgtattccagcagcaatagggg 2940
QY 2941 CCTGCTATTGCTAAAGAAATAAGGTCCAGACGAAAGGAGTGTATGTTGAACTTAAATGTCG 3000
DB 2941 cctgctattgcttaagaaataaaggtccagacgaaaggagtgatcaggcctctgcaacgg 3000
QY 3001 AAATCTAGCTTCCAGCATTTGAGATCTTGGAGTCTCGGGCATCAAGCTGGTTGAGAAAG 3060
DB 3001 aaatctagcttccagcatcttgagatctcttgagctcgctgggcatcaagctggtgagaaag 3060
QY 3061 TGGAGACTTGAAGCTTCCCTGATACACAGATTAAGATAGTATATAGCAGGGCACATG 3120
DB 3061 tggagacttgaagcttccctgatatacagataaagatagtagatgatatagcagggacatg 3120
QY 3121 TATCTTTGTATTAACTCCGTTCTGGAATATATATTTGTAACCTTAAATGTCGACAAATA 3180
DB 3121 tatctttgtattaaactccgttctggaatatatatatttgtaactaaaatgtgacaataa 3180
QY 3181 AAAGAACGGGTGGAGTATATTTGAAGAGCGGCAAGAAACCTCTGTATATATGACCTGT 3240
DB 3181 aaagaaacgggtggagtatatttgaagagcggaagaaacacctctgtatataatgacctgt 3240
QY 3241 CGATATCAATAATAATGCCGATCAGTT 3265
DB 3241 cgatatcaataataatgccgatcagtt 3265

RESULT 2

AAV99564

ID AAV99564 standard; cDNA to mRNA; 1908 BP.

AC AAV99564;

DT 29-MAR-1999 (first entry)

DE Maize lysine ketoglutarate reductase cDNA.

KW lysine ketoglutarate reductase; saccharopine dehydrogenase;

KW transgenic plant; seed; soybean; ss.

OS Zea mays.

FH Key Location/Qualifiers

FT CDS 3..1908

FT /*tag= a

XX

PN W09842831-A2.

XX

PD 01-OCT-1998.

XX

PF 27-MAR-1998; 98WO-US06051.

XX

PR 27-MAR-1997; 97US-0824627.

XX

PA (DUPO) DU PONT DE NEMOURS & CO E I.


```
Db 1501 ttattgattccttgactctcttagcttaataagacatggtggagatcacgagtcgcggcgaag 1560
QY 1561 AAATTGAATTAGCTCTGACAGATAGGAAAAGTCAATCAGTATGAAACTGACGTACAAATTTG 1620
Db 1561 aaattgaattagctctgaagataggaagaaagtcgaatgagatgaactgaactgcacaaattg 1620
QY 1621 ATAAGGAGGGCCAAAGATTTTAAATCTTGGAGCTGGAGAGATCTGTGGCCAGCTGCTG 1680
Db 1621 ataaaggaggccaaagatttaattctctggagctgggaagagtcgtgcggcagctgctg 1680
QY 1681 AGTTTCTGCATCTTACCAGACATATGATGACCTGATGACCATGATGACAGATC 1740
Db 1681 agttctgcgcatctaccagacatgtacctatggtgtgtagaccatgatgcagatc 1740
QY 1741 AAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCGAAGAGACACATTTGATGTA 1800
Db 1741 aaattcattgtatctggtgcatctttgtatcaaaaagatgcagaagagacagattgatgta 1800
QY 1801 TTCAAAATACACTGCTACCCAGCTTGATGTTGCTGATATTTGGAGCCTTTTCAGATCTTG 1860
Db 1801 tcyaaaatacaactgctacccagcttgatgttgcgtgatattggagcctttccagatcttg 1860
QY 1861 TTTCTCAGGTTGAGGTTGATTAATAGCTTGCTGCTGCTGCTAGTTTTCATG 1908
Db 1861 ttctcaggttgaggttgtaattagctgctgctgctgctgctgctgctgctgctgctgctg 1908

RESULT 3
AC42953
ID AAC42953 standard; DNA; 3195 BP.
XX
AC AAC42953;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37480.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PF 06-SEP-2000.
XX
PP 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 28-APR-1999; 99US-0130510.
PR 30-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 05-MAY-1999; 99US-0132407.
PR 06-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 23-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0136029.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
```

PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	26-OCT-1999;	99US-0161320.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 37.8%; Score 1233.2; DB 21; Length 3195;			
Best Local Similarity 64.5%; Pred. No. 0;			
Matches 1997; Conservative 1; Mismatches 1029; Indels 67; Gaps			
QY	15	CTGCTAGGAGGAGCAGACGGACCTCGAGTAAACCCGATATTGTGCAGCCAAGCAC	74
Db	121	ctttLacagctgggaaagacagacgcggcattccccgcattgtggttcagccatcgtc	180
QY	75	AGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCCAGATTCCTGAGAAGACCTG	134
Db	181	adgcgtatccatcaltgatgccttgtatgaagatgttgggtgtaaaattttctgatgatgtt	240
QY	135	TCAGAAATCGGCGCTTATCATFAGGATCAACACACCCCAAGCTGCAGATGATTCCTTCAGAT	194
Db	241	tctgatgtggccttatacttggtaaccaaacocctgagctagaagaatgattctccagag	300
QY	195	AGAGCGTACGCTTCTCTTTACACACACACAAAGCCCAAAAGAGAAATATGCCACTGTTA	254
Db	301	agagcatagcgtttctttcacatactcataagggcacagaagagacatgcctttgttg	360
QY	255	GACAAAGATCCTTGAGAAAGGCTGCTTGTTCATTTATGAGCTAATGTGTGGAGATGAT	314
Db	361	gataaaattctctgagagagtgactttgttgattatgagctcatgtttggggatcat	420
QY	315	GGGAAAGATCACTAGCATTTGGGAAATTTGCTGCTAGAGCTGCAGCTAGATFAGATTTCTTA	374
Db	421	gggaacagattattggcggtttggtaaataatgcaggagagctggctgtgttgactttct	480
QY	375	CATGGCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCATTTCTCTCTCGGA	434
Db	481	cacggacttgacagcgtgaagctcatctcaggactactcaacacattctcctcgctcgt	540
QY	435	CAATCTCATATGTATCTCTCGTCTGCTGAGCCCAAGGCTGCAGTCAATTCGTTGCAGAA	494
Db	541	gcaltgatattgtattcctcatgttgctgctgcaaaagcgcgtLaatttctgttggtag	600
QY	495	GAGATAGCAACATTTGGACTTCATCCGAAATTTGTCGAGTAGTGTGTGTTCATCTGGA	554
Db	601	gaaattgcaagccaggggaactgccttaggaaactcgcctctgttattgttcttcaccgga	660
QY	555	GTGTGAAACGCTCTCAGGCTGGCAGGAGATATCAAGTATTATGCCCATACCTTTGTT	614
Db	661	acaggaaaatgttctcttgggggcgaagaaattttcaagcttcttcacactttgtt	720
QY	615	GATGCTGAGAGAGCTTCGGAATTTTTCAGGCCAGGAATCTGCTFAAGCAATCTCAG---	671
Db	721	gaaccaagcaactctctgaactatttgttaaaagacaagaagaattagtcacaaatgggatt	780
QY	672	TCGACCAAGAGAGTATTCAACTTTTATGGTTGTGTGACTCTFAGAGACATPATTCT	731
Db	781	tcacaagaagcagtcatacgaatattggttatattaccagccaagacatggttgaa	840
QY	732	CACAAGGATCCCACAGACAAATTTGACAAAGGTGACTATTATGCTCATCCGACACATC	791
Db	841	cacaagaLccatcaagctcatcgaagaagcgcactattatgcacaccggaacattac	900
QY	792	ACCCCTGTTTTCATGAAGAATTTGCTCCATATGCATCTGTCATCGTAACCTGATGAT	851
Db	901	aatccagttttccagcaaaagatatcgcataacgtctgttctgttaaatgtgtgac	960
QY	852	TGGGAGAGAGGTTTCCACCATTTACTAAATATGGATCAGTTTACAGCAATTTGATGGAGCT	911

Query Match	37.8%;	Score 1233.2;	DB 21;	Length 3195;	
Best Local Similarity	64.5%;	Pred. No. 0;	Mismatches 1029;	Indels 67;	Gaps
Matches 1997;	Conservative	1;			
PR 23-JUL-1999;	99US-0145224.				
PR 26-JUL-1999;	99US-0145276.				
PR 27-JUL-1999;	99US-0145913.				
PR 27-JUL-1999;	99US-0145918.				
PR 27-JUL-1999;	99US-0145919.				
PR 28-JUL-1999;	99US-0145951.				
PR 02-AUG-1999;	99US-0146386.				
PR 02-AUG-1999;	99US-0146388.				
PR 03-AUG-1999;	99US-0146389.				
PR 03-AUG-1999;	99US-0147038.				
PR 04-AUG-1999;	99US-0147204.				
PR 04-AUG-1999;	99US-0147302.				
PR 05-AUG-1999;	99US-0147192.				
PR 05-AUG-1999;	99US-0147260.				
PR 06-AUG-1999;	99US-0147303.				
PR 06-AUG-1999;	99US-0147416.				
PR 09-AUG-1999;	99US-0147493.				
PR 09-AUG-1999;	99US-0147935.				
PR 10-AUG-1999;	99US-0148171.				
PR 11-AUG-1999;	99US-0148319.				
PR 12-AUG-1999;	99US-0148341.				
PR 13-AUG-1999;	99US-0148565.				
PR 13-AUG-1999;	99US-0148684.				
PR 16-AUG-1999;	99US-0149368.				
PR 17-AUG-1999;	99US-0149175.				
PR 18-AUG-1999;	99US-0149426.				
PR 20-AUG-1999;	99US-0149722.				
PR 20-AUG-1999;	99US-0149723.				
PR 20-AUG-1999;	99US-0149929.				
PR 23-AUG-1999;	99US-0149902.				
PR 23-AUG-1999;	99US-0149930.				
PR 25-AUG-1999;	99US-0150566.				
PR 26-AUG-1999;	99US-0150884.				
PR 26-AUG-1999;	99US-0151065.				
PR 27-AUG-1999;	99US-0151066.				
PR 27-AUG-1999;	99US-0151080.				
PR 30-AUG-1999;	99US-0151303.				
PR 31-AUG-1999;	99US-0151438.				
PR 01-SEP-1999;	99US-0151930.				
PR 07-SEP-1999;	99US-0152363.				
PR 10-SEP-1999;	99US-0153070.				
PR 13-SEP-1999;	99US-0153758.				
PR 15-SEP-1999;	99US-0154018.				
PR 20-SEP-1999;	99US-0154039.				
PR 22-SEP-1999;	99US-0154779.				
PR 23-SEP-1999;	99US-0155139.				
PR 23-SEP-1999;	99US-0155486.				
PR 24-SEP-1999;	99US-0155659.				
PR 28-SEP-1999;	99US-0156458.				
PR 29-SEP-1999;	99US-0156596.				
PR 04-OCT-1999;	99US-0157117.				
PR 06-OCT-1999;	99US-0157753.				
PR 06-OCT-1999;	99US-0157865.				
PR 08-OCT-1999;	99US-0158023.				
PR 12-OCT-1999;	99US-0158369.				
PR 12-OCT-1999;	99US-0159283.				
PR 13-OCT-1999;	99US-0159294.				
PR 13-OCT-1999;	99US-0159295.				
PR 14-OCT-1999;	99US-0159329.				
PR 14-OCT-1999;	99US-0159330.				
PR 14-OCT-1999;	99US-0159331.				
PR 14-OCT-1999;	99US-0159637.				
PR 14-OCT-1999;	99US-0159638.				
PR 18-OCT-1999;	99US-0159584.				

Tue Jun 4 15:23:05 2002

XX Sequence 2582 BP; 798 A; 482 C; 562 G; 739 T; 1 other;
 SQ
 Query Match 27.8%; Score 907; DB 20; Length 2582;
 Best Local Similarity 62.7%; Pred. NO. 4.2e-247; Indels 51; Gaps 6;
 Matches 1531; Conservative 2; Mismatches 857;
 QY 727 TTCTCACAAGAGTCCACACAGCAATTTGACAAAGGTGACTATTATCTCATCTCCTCACTGTA 786
 Db 1 ttgaacccaaatcactgtagtggttgacaaagcagactactattcacacctgagc 60
 QY 787 ACTACACCCCTGTTTTCATGAAGAATTCCTCATATGCTCTCTCATCTGTA 846
 Db 61 attacaatcccactttccatgataaataagcaccatgcatctgtattgtcaattgca 120
 QY 847 TGATATGGGAGAGAGGTTTCCACCACTTACTAAATATGATGATGATGATGATGATG 906
 Db 121 tttatggggagaaagatttcccaattgcccagactataagcagatgcaagacttaattg 180
 QY 907 AGACTGGTTCTCTTGTAGTCGGCGTTTGTGACATACACTTGTGATATGAGGTTCATTG 966
 Db 181 gccggggagagcccccctgttggaatagctgacataacgtgtgatacaggggttcaattg 240
 QY 967 AATTTATCAACAAGAGTACATCAATAGAGAGCGCTTCTTTCGGTATGATGATGATGATG 1026
 Db 241 agttgttaaccgggacttccaaatgattcacccttctcagatgataccttcaacaa 300
 QY 1027 ATTCATACCATGATGATGATGAGGTCGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1086
 Db 301 attcctaccatgatatggagggaatggagtgatgctgagctgattgctgagcttcttc 360
 QY 1087 CTACAGAATTCCTTAAGAGGCTCCCAACATTTTGAAACATACATCTATCTAGACTTGTG 1146
 Db 361 caacagaatttgcgaaggaggtcccaacattttgaaacatacttcccaattgttg 420
 QY 1147 CTAGTTTGGCTCAGTGAAGCAACCGGCAACTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1206
 Db 421 taaattggctctgctacagacattacaagttgctcacttaagagaggtctgca 480
 QY 1207 TTGCACATGCTGGCAGATTAATCTCTTGTATGAATATATATCTCTCTCTCTCTCTCTCT 1260
 Db 481 tagccataaaggagtgctacacctcttatgattatccacgcagctggaggtctg 540
 QY 1261 ATACTATGATAGATTTGGCACCCGCCCAAAATCCATTCCTCTCTCTCTCTCTCTCTCTCT 1320
 Db 541 attcagaggagatcagaaaaacgcagaaaaattctctatcccaaaaaaggaggtacata 600
 QY 1321 CCTGTATCTCTCAGTGGGCACCTATTGTAAGTATCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
 Db 601 tctcgtgctctgagtgagtgctacttattgacagttctgataaatgagccttagata 660
 QY 1381 TCATTGAGACAGCTGGAGGTTTCATTTCACCTGGTTAGATGTGAAGTTCGACAAAGCAGG 1440
 Db 661 ttattgagtgaggaggtctctcacttagtcaactgcaactgcaactgcaactgcaactgca 720
 QY 1441 ATGATATGATCTCAGAGCTTGAAGTAGAGAGAGATGATGATGATGATGATGATGATGATG 1500
 Db 721 aagccgtatctctctgacactgaaagtggagagagagagagagagagagagagagagag 780
 QY 1501 TTATTGATTCCTTGACTCTTTTGTAGCTTAATGACATGCTGGATGATGATGATGATGATGAT 1560
 Db 781 tcatgattctttaaactgctattgctagtcacactgaacatgataatttcaaatcaag 840
 QY 1561 A---AATTGAATGACTCTGAAGATAGGAAAGTCAATGATGATGATGATGATGATGATGATGAT 1617
 Db 841 attcaagataaatttcaactaaagcttgtaagttaagttaagttaagttaagttaagttaagt 900
 QY 1618 TTGA---TAAAGGAGGCGCCCAAGATTTAATTTCTGGAGCTGGAAGAGTTCGCGCCGAG 1674
 Db 901 ctgacccccagaaagaggtgcggttttaattcttgagctgctgaggtctgctcaaccag 960
 QY 1675 CTGCTGAGTTTCTGCGATCTTACCCAGACATATGATGATGATGATGATGATGATGATGATGAT 1713

QY 3031 AGTCCTCGGCAATCAAGCTGGTTGAGAAAGTGA 3064
 Db 3158 aagcatatggtataaagctgtagtgagaagggcaga 3191
 RESULT 4
 AAV99562
 ID AAV99562 standard; cDNA to mRNA; 2582 BP.
 XX AAV99562;
 AC
 XX 29-MAR-1999 (first entry)
 DT
 XX Soybean lysine ketoglutarate reductase cDNA.
 DE
 XX Lysine ketoglutarate reductase; saccharopine dehydrogenase;
 KW transgenic plant; seed; soybean; ss.
 KW
 XX Glycine max.
 OS
 XX Key Location/Qualifiers
 FH CDS 3..2357
 FT /*Cag= a
 FT
 FN W09842831-A2.
 PN
 XX 01-OCT-1998.
 PD
 XX 27-MAR-1998; 98WO-US06051.
 PF
 XX 27-MAR-1997; 97US-0824627.
 PR
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 PA
 XX Epelbaum SU, Falco SC, McDevitt RE;
 PI
 XX WPI: 1999-045139/04.
 DR P-PSDB; AAW87760.
 XX
 XX Nucleic acids and chimeric genes for increasing seed lysine content
 PT - comprise sequence encoding all or part of lysine ketoglutarate
 PT reductase, useful to improve nutritional quality of seeds from
 PT transformed plants
 PT
 XX Claim 3; Page 182-184; 231pp; English.
 PS
 XX This is the nucleotide sequence of a near full-length cDNA encoding
 CC soybean lysine ketoglutarate reductase (LKR, see AAW87760), or
 CC saccharopine dehydrogenase (SDH). It was obtained by PCR
 CC amplification of developing seed cDNA using primers (see AAV99560-61)
 CC based on Arabidopsis LKR/SDH sequences, and by RACE. Isolated
 CC nucleic acids comprising sequences encoding all or part of plant
 CC LKRs are new. Also claimed are: (1) a chimeric gene comprising the
 CC fragment (or a subfragment) operably linked to a seed specific
 CC regulator, where the chimeric gene reduces LKR activity in plant
 CC seeds transformed with it; (2) plant cells and seeds in which LKR
 CC activity is reduced due to a mutation in the gene encoding LKR or
 CC transformation with the chimeric gene; (3) a nucleic acid fragment
 CC comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
 CC in which a nucleic acid fragment encoding dihydrodipicolinic acid
 CC synthase (DHDPS) substantially insensitive to lysine inhibition is
 CC operably linked to a plant chloroplast transit sequence and to a
 CC plant seed-specific regulatory sequence; (4) plants comprising in
 CC genome (i) and (ii), especially as fragment of (2); and (5) seeds
 CC from (3). The chimeric genes can be used to produce plant cells
 CC and seeds with reduced LKR activity, especially in Arabidopsis,
 CC corn, soybean, rapeseed, wheat and rice (claimed). LKR is
 CC important in controlling free lysine accumulation in plant seeds.
 CC LKR activity reduction may be achieved by cloning the claimed
 CC fragment, preparing a chimeric gene for cosuppression of LKR,
 CC expression of antisense RNA for LKR, and transforming plants with
 CC the chimeric gene.

Db 961 ctgctgaagtgtatcatcatcttgggaagccatcatcagcaaatggtataaaacattgt 1020
QY 1714 ATGGTGTGATGACCATGATGAGATCAAAATTCATGTTATGCTGTCATCTTTGATCAAA 1773
Db 1021 tgggaagatgatttgaattgcaactgctgtagaagtcattgtggatctctgctacctga 1080
QY 1774 AAGATGCAGAGACAGCTTGATGTTATGAAATACAACTGCTACACAGCTTGATGTTG 1833
Db 1081 aggaatgcagacagactgtgagggcattcccaaatgtaaccggcaattcagctgatga 1140
QY 1834 CTGATATGGAGCCCTTCAGATCTGTTCTCAGGTGAGTGTGTAATAGCTTGCTGC 1893
Db 1141 tggatcggtgccaatttgttaaglacatttcacaggttgacgttggatagtttgcgc 1200
QY 1894 CTGCTAGTGTTCATGCTGCCATTCAGGAGTATGCATAGAGTGTGAAGAAGCATGGTAA 1953
Db 1201 ccccaagttgcataattatgtagcaaatgctgcatgagctggaataaacatctgtga 1260
QY 1954 CGGCAGCTATGTTGATGATFCCATGTCATGTCAACTTGAAGCTGCAAGCTGCAAGTG 2013
Db 1261 ctgtagctatgttgatagctccatgctcaatgctcaatgataaggtctaaagatgcggca 1320
QY 2014 TAACTATACTTGTGAAATGGCTAGATCTCGCATAGATCACTTGATGTCAATGAAGA 2073
Db 1321 taacaattcttggagagatggcttgagccaggaattggtcatatgatggcaatgaaga 1380
QY 2074 TGATGTGATGAGCTCATGACGAAAGGAAATAAAGGCATTTACATCTTACTCTGCTG 2133
Db 1381 tgaatcaaccaagcacatgtgaggaagggaataaagcttcctcaactcttcttattggtg 1440
QY 2134 GATTGCCATCTCCAGCTGACGAAACAATCCGCTTCCTGCTATAAATTCAGTGGAAACCCAG 2193
Db 1441 gacttccatctctgaactgctgcaacaatccattagcatataaattcagttggaatcctg 1500
QY 2194 CTGCTGACTCCGCTCAGGGAATAATCTGCAGTCTACAAATTTCTGTGTGAGAGCATCC 2253
Db 1501 caggagccatccgagctggcgcaactcctgcacactacaatggggtggtgaaactgtac 1560
QY 2254 ATGTAGATGGTCAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2313
Db 1561 atattgatgggagacatcttattgatgctggtacaaagactaaggctacccggaacttctg 1620
QY 2314 CTTTGTCTGGAACACTTGCACAAATCGAAATCTCTGATATGATGATGATGATGATGATG 2373
Db 1621 ctttggcttggaaatgctcccaaatcgcaaatcttacttacttattggtggtggtgaa 1680
QY 2374 TCTCAAAAGCAAGCATCCACCATATATAGGGTACTTTCGTTTACGAAGTTTATGATGAG 2433
Db 1681 t---aactgaagacaaaccatttccgtggaacccctcgcctatgaaggatttagtga 1737
QY 2434 TTATGGTAACCTTTCCAAAATCGGTTCTTTGATGCTGCAAAATCATCCACTGCTGCAAG 2493
Db 1738 teatgggacactgctaggtattgatttatacaatgaagccatctcgttctgaatga 1797
QY 2494 ATACTAGTCTGCAACATATAGGGTTTCTTGTGATGATGATGATGATGATGATGATGATG 2553
Db 1798 atggcaaaagcaactctcaaaaattcttattgaaacttctcaaatggttgggtgata 1857
QY 2554 TTACACGGACTTATAGATATTGAAGCTTCTGTGTGATGATGATGATGATGATGATGATG 2613
Db 1858 atccagatgaactattgatga-----gagaatgacatgagagcaaa 1902
QY 2614 TGTGTAAGCTCGGTGTGCAAAATTAAGAAATAGCTGTTAAGACAGCTCAAAACCATCA 2673
Db 1903 tattaatacaagggcactgcaagatcaaaagacggcaatgggagcagcaaaaacaatca 1962
QY 2674 AGTTCTTGGACATACATGAAGAGACTCAATACCTAAGGTTGTTCCAGCCCATTTGATG 2733
Db 1963 ttttcttgggacttcttgacaaactgaaatccctcttccctgcaaaagctgttttgatg 2022
QY 2734 TGATTTGCCAGCAATGGAAACAGAGATGCCCTATGGCCACAAATGAGCAAGACATGGTAC 2793

Db 2023 ttgctgttttcgcgcattgaggagaggttatcatcacccagcaagaaagatatggtgc 2082
QY 2794 TGCTCACACCAAGTCGAGGTGGAATACCCGCGGCAACCCGCCGCAAAAGCACCAG 2853
Db 2083 ttttgcattcatgagtggaatagaaatagaaatcccaagatagcaaatctacagagagcatag 2142
QY 2854 CGACGCTACTGAGTTCGGGAAGGTTGAAATGGCAGGTCCACCACCTGCCATGGCGTGA 2913
Db 2143 ctacttacttgaatttgggaagactcttgataaaaaaacacacactgcctgccccctta 2202
QY 2914 CGCTCGCATTCACAGCAATAGGGCCCTTGCTTAAAGTAAGTTCAGATTCAGACGA 2973
Db 2203 ctggttggtattccagctgctgttgagagcttggagcttggcttttattgacaaaattcagacaa 2262
QY 2974 AAGGAGTATGATGAGGCTCTGCAACCGGAAATCTACGTTCACAGCATTTGGAGT 3033
Db 2263 gaggagcttcttaagcctatcgaaactgaagtatacaatccagcactggatattatagaag 2322
QY 3034 COTCGGCATCAAGCTGTTGAGAACTGGAGACTTGAAGTTCCTGATACACAGATAA 3093
Db 2323 ctatggggatcaagttgatagagaagaccgagtaatttgcattatgaattgatgtatg 2382
QY 3094 AGATAGTATGATATAGCAGGCAATGATATCTTTTGTATTA 3134
Db 2383 gtgcacattatgtacacccatgcaatgttggatttgaataa 2423

RESULT 5

AAV99565
ID AAV99565 standard; cDNA to mRNA; 720 BP.
AC AAV99565;
XX AAV99565;
DT 29-MAR-1999 (first entry)
XX Rice lysine ketoglutarate reductase cDNA.
DE Lysine ketoglutarate reductase; saccharopine dehydrogenase;
XX transgenic plant; seed; rice; ss.
KW Oryza sativa.
XX
FH Key Location/Qualifiers
FT CDS 2..720
FT /*tag= a
XX
PN WO9842831-A2.
XX
PD 01-OCT-1998.
XX
PF 27-MAR-1998; 98WO-US06051.
XX
PR 27-MAR-1997; 97US-0824627.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Epelbaum SU, Falco SC, McDevitt RE;
XX
DR WPI; 1999-045139/04.
XX
P-PSDB; AAW87763.
XX

Nucleic acids and chimeric genes for increasing seed lysine content
- comprise sequence encoding all or part of lysine ketoglutarate
reductase, useful to improve nutritional quality of seeds from
transformed plants

Claim 3; Page 196; 231pp; English.

This is the nucleotide sequence of a partial cDNA encoding a
rice lysine ketoglutarate reductase (LKR) or saccharopine
dehydrogenase (SDH) polypeptide (see AAW87763). It was isolated
from an expressed sequence tag database by comparison to an
Arabidopsis LKR/SDH protein (see AAW87759). Isolated nucleic

acid fragments comprising sequences encoding all or part of plant LKRs are new. Also claimed are: (1) a chimeric gene comprising the fragment (or a subfragment) operably linked to a seed specific regulator, where the chimeric gene reduces LKR activity in plant seeds transformed with it; (2) plant cells and seeds in which LKR activity is reduced due to a mutation in the gene encoding LKR or transformation with the chimeric gene; (3) a nucleic acid fragment comprising: (i) chimeric gene above, and (ii) a second chimeric acid in which a nucleic acid fragment encoding dihydrodipicolinic acid synthase (DHDDS) substantially insensitive to lysine inhibition is operably linked to a plant chloroplast transit sequence and to a plant seed-specific regulatory sequence; (4) plants comprising in genome (i) and (ii), especially as fragment of (2); and (5) seeds from (3). The chimeric genes can be used to produce plant cells and seeds with reduced LKR activity, especially in Arabidopsis, corn, soybean, rapeseed, wheat and rice (claimed). LKR is important in controlling free lysine accumulation in plant seeds. LKR activity reduction may be achieved by cloning the claimed fragment, preparing a chimeric gene for cosuppression of LKR, expression of antisense RNA for LKR, and transforming plants with the chimeric gene.

XX Sequence 720 BP; 203 A; 150 C; 158 G; 206 T; 3 other;

Query Match 16.0%; Score 522.6; DB 20; Length 720;
Best Local Similarity 85.5%; Pred. No. 4.7e-138;
Matches 579; Conservative 1; Mismatches 97; Indels 0; Gaps 0;

QY 1848 CTTTCAGATCTTTCTCAGGTTGAGTGTGTAATAGCTGTGCTGTCTAGTTTTCAT 1907
DB 11 ctttccaaactgtttctcagggttgaagtagtagttagctgtcgtcgcaggtttcat 70

QY 1908 GCTGCCATTCGAGGATGTCATAGATTGAAGACACATGTTACGCGCAACTATGTT 1967
DB 71 gctgcacgaagagtagtgcataagatgaagagcacttggctcgtcgaagctatggt 130

QY 1968 GATGAATCCATCTCAAACTTCAGCCAGCTGCCAAAGATCAGGTGTAACATACATTTGT 2027
DB 131 gatgagtcacagtgcaagagttggaacaactctgcagaaggtgctggtgaactattctctg 190

QY 2028 GAAATGGGCTAGATCTCGGATAGATCATTGATGTCAATGAGATGATGATGAGCT 2087
DB 191 gaaatgggctggtacccctggcatanattcatatgatgtcgaatgaagatgattgcgaagca 250

QY 2088 CATGCAGAAAGGAAATAAAGCAATTAACATCTTACTGTGGTGGATGCCATTCCTCA 2147
DB 251 catcgcaggaagggaataaagtcattacatctcttggagggaactcccatctcca 310

QY 2148 GCTGCAGCAAAACATCCGCTTCCTTATAATTCAGTTGGACCCAGCTGGTGCCTCGG 2207
DB 311 gcttcgcaacaacatccactgtcttataagttcagttcagttcagttcagttcagttcagtt 370

QY 2208 TCAGGAAATAATCTCGACTCTACAAATTTCTGTTGAGAGCATCCATGTAGATGGTCAAT 2267
DB 371 gcaggagaaacccctgctctcacaattctcgaagaaatccatccatcgtatgagtgtgat 430

QY 2368 AACTTGTATCAATCAGCAAGAGGCTCAGACTACGAGAGCTTCAGCTTTTCTCTGAA 2327
DB 431 aaattgtatcaatccgcaagaggtccagattacmagaactccagcttccagcttccagcttcc 490

QY 2328 CACTTGCCTCAATCGGAATTCCTTGATATATGTTGAGCTTTATGGTATCTTCCAAAGAGCA 2387
DB 491 cacttgcctcaacgggaactcctctgctcagagacactgtatgggattctcccaagaagca 550

QY 2388 TCCACCATATATAGGGCTACTTTCGTTACGAGGTTTATGAGATTTATGTATACCCCTT 2447
DB 551 tctactgtacagggctactctctgttatgaagagtttaataatgaagataatggaacacttc 610

QY 2448 TCCAAACTGGGTTCTTTGATGCTGCAAAATCATCCATCTGCTGCAAGATACTAGTCTCCA 2507
DB 611 gcgaaattgggtttttgtatgctgcaggtcattccactgttgcaacaactactgcacct 670

QY 2508 ACATATAAGGTTTCTCT 2524
DB 671 acatacangatttct 687

RESULT 6
AAV99568
ID AAV99568 standard; DNA: 1449 BP.
XX AAV99568;
AC AAV99568;
XX 29-MAR-1999 (first entry)
DT 29-MAR-1999 (first entry)
XX Arabidopsis saccharopine dehydrogenase coding region.
DE Arabidopsis saccharopine dehydrogenase;
XX lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW transgenic plant; seed; ds.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN WO9842831-A2.
XX 01-OCT-1998.
PD 01-OCT-1998.
XX 27-MAR-1998; 98WO-US06051.
PF 27-MAR-1997; 97US-0824627.
PR 27-MAR-1997; 97US-0824627.
XX (DUPO) DU PONT DE NEMOURS & CO E I.
PA Epelbaum SU, Falco SC, McDevitt RE;
PI WPI; 1999-045139/04.
XX P-PSDB; AAW87766.
DR Nucleic acids and chimeric genes for increasing seed lysine content
XX - comprise sequence encoding all or part of lysine ketoglutarate
PT reductase, useful to improve nutritional quality of seeds from
PT transformed plants
XX Claim 3; Page 200; 231pp; English.

This nucleotide sequence of a coding region for an Arabidopsis thaliana polypeptide (see AAW87766) comprising the saccharopine dehydrogenase (SDH) domain of the lysine ketoglutarate reductase/SDH protein. Isolated nucleic acid fragments comprising sequences encoding all or part of plant LKR polypeptides are new. Also claimed are: (1) a chimeric gene comprising the fragment (or a subfragment) operably linked to a suitable seed specific regulator, where the chimeric gene reduces LKR activity in plant seeds transformed with it; (2) plant cells and seeds in which LKR or activity is reduced due to a mutation in the gene encoding LKR or transformation with the chimeric gene; (3) a nucleic acid fragment comprising: (i) chimeric gene above, and (ii) a second chimeric gene in which a nucleic acid fragment encoding dihydrodipicolinic acid synthase (DHDDS) substantially insensitive to lysine inhibition is operably linked to a plant chloroplast transit sequence and to a plant seed-specific regulatory sequence; (4) plants comprising in genome (i) and (ii), especially as fragment of (2); and (5) seeds from (3). The chimeric genes can be used to produce plant cells and seeds with reduced LKR activity, especially in Arabidopsis, corn, soybean, rapeseed, wheat and rice (claimed). LKR is important in controlling free lysine accumulation in plant seeds. LKR activity reduction may be achieved by cloning the claimed fragment, preparing a chimeric gene for cosuppression of LKR, expression of antisense RNA for LKR, and transforming plants with the chimeric gene.

Sequence 1449 BP; 446 A; 267 C; 356 G; 380 T; 0 other;
Query Match 15.8%; Score 515; DB 20; Length 1449;
Best Local Similarity 61.7%; Pred. No. 1e-135;

Db 505 tgagcgaatcatcgacgaacgcgagagacagcaggtggcctttggcaaatatgcccggagt 564
Qy 353 ACCTGGAGCTGATAGATTCTTACATGCTTCGGACAGCGGATATTTGAGCGTTGGATACATC 412
 ||||| | | | | | | | | | | | | | |
Db 565 ggcggcatggtgaacatcctgcacgcatggtgactgcgtcttttggccctgggacatca 624
Qy 413 GACTCCATTTCTCTCTGGGACAACTCATATGATATCTTCCTCGCTCGCAGCCAAAGGC 472
 ||||| | | | | | | | | | | | | | |
Db 625 tacaccattcatgacattggaccctgcataactatcgcaattctctcgatggctcgcca 684
Qy 473 TGCAGTCATTGCTGTCGACGAAGAGATAGCAACATTTGGACTTCCATCCGGAAATTTGTCC 532
 ||||| | | | | | | | | | | | | | |
Db 685 ggcattccggattggtgctacagatctcgttggcatgattgcccgaagtcctattggacc 744
Qy 533 GATAGTGTGTTGTTCTACTGAGTGTGAAACGCTCTCAGGGTGGCGCAGGAGATATTCAA 592
 ||||| | | | | | | | | | | | | | |
Db 745 acttacattgtgtctactggtcgtggaatgtttcccaaggtgcccagaggtgttctc 804
Qy 593 GTTATGCCCATACCTCTTGTGATGCTGAGAGCTTCCCGAAATTTTTCAGGCCAGGAA 652
 ||||| | | | | | | | | | | | | | |
Db 805 cgagctgccattgagtattgtccaccgagagatgctacgaagtggtggcgaacatggaaa 864
Qy 653 TCTGFTTAAGCAATCTCAGTCGACCAAGAGAGATTTCAACTTTATGTTGTTGTGTAC 712
 ||||| | | | | | | | | | | | | | |
Db 865 tcaaaataagctttlacggtcgcgaggtgagccgatcgatcatctgagcgtcgtgaggg 924
Qy 713 CTCATAGACATAGTTTCTCAACAGGATCCACCAGACAAATTTGACAAAGGTGACTATTA 772
 ||||| | | | | | | | | | | | | | |
Db 925 cggcggatttatgcccgaaggaatcagatgagttcccgcgaacgatacatctccaccttag 984
Qy 773 TGCTCATCAAGCAACTACACCCCTGTTTTCATGAAGAATTTGCTCCATATGATGATCTGT 832
 ||||| | | | | | | | | | | | | | |
Db 985 caacgaatagctcctgcacgcatcgtttattgttaacgycatctactggcgttaggcag 1044
Qy 833 CATCTGAACCTATGTTATGGGAGAGAGGTTTCCACCATTACTAAATATGGATCAGTT 892
 ||||| | | | | | | | | | | | | | |
Db 1045 tcccaagttgatcagcatctccgagtcgcaagaatctgctacgtccggcaaacactccctg 1104
Qy 893 ACAGCAATTTGAGGAGCTGTTGCTCCTTTAG-----TCGGCGTTTGTGACAT 940
 ||||| | | | | | | | | | | | | | |
Db 1105 gctgcgcgacgaaggttagtccggctttgcccacatcgcatgctggccatttgcgacat 1164
Qy 941 ACATTGTGATATGAGGTTCCATGAATTTATCAACAAGAGTACATCAATAGAGAGGCC 1000
 ||||| | | | | | | | | | | | | | |
Db 1165 ttccgctgatcccggtggtccattgagtttatgaacgagtgccaccaccatcgacatcc 1224
Qy 1001 TTTCTTTCGGTATGATCTCTTAAGATTTTATACCATGATGATATGGAAGGTGCCGAGT 1060
 ||||| | | | | | | | | | | | | | |
Db 1225 attctgctgtacgacgcgacaggaataaggatacccaagagctttaaggagaccggagt 1284
Qy 1061 GGTCTGCTGGCTGTGACATCTCCCTACAGAAATCTCTTAAGAGGCTCCCAACATTT 1120
 ||||| | | | | | | | | | | | | | |
Db 1285 tttagtctgttcgattgataatgctcaccaattgcccagagaatcgacggatttggtt 1344
Qy 1121 TGGAAACATATCTATCTAGACTTGT 1145
 ||||| | | | | | | | | | | | | | |
Db 1345 tggagagctgttggctcctcatgtt 1369

RESULT 10
ABL12396/C
ID ABL12396 standard; cDNA; 6209 BP.
XX
AC ABL12396;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31670.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.

XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX PA Venter JC, Adams M, Li PWD, Myers EW;
XX PI WPI; 2001-656860/75.
XX DR P-PSDB; ABB68293.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Claim 1; SEQ ID NO 31670; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (AB857737-AB872072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 6209 BP; 1742 A; 1402 C; 1329 G; 1736 T; 0 other;

Query Match 3.4%; Score 110.4; DB 23; Length 6209;
Best Local Similarity 49.7%; Pred No. 2.9e-20;
Matches 310; Conservative 0; Mismatches 311; Indels 3; Gaps 1;
QY 1868 GGTGTGAGTGTGTAATAGCTTCTGCTGCTGCTAGTTTTCATGCTGCATGAGGAGTATG 1927
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2259 GGCAGATGTAGTAGTTTCCCTGCTGCTGCTACAGTCTCCATGCTGCGACGTTATG 2200
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1928 CATAGAGTTGAAGACGACATGTAACGCAAGCTATGTTGATGAATCCATGTCATAACT 1987
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2199 CGTGGCAGAGGGCCACCACATGTTACCCCAAGTTATCTGAACGACGAGATCTCCGGTT 2140
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1988 GAGCCAAAGCTGCCAAAGATGCAGGTGTAATACTATCTTTGTGAATGGCGCTAGATCCCTG 2047
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2139 GCACGAGGAGGCCAAGGCCAAGGCTGTGACCATCATGAACGAGGTGGCTTGGATCTCTGG 2080
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2048 CATAGATCATCTTATGTCATGAAGATGATGATGAAGCTCATGCACGAAAGGAAAAAT 2107
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2079 TATCGATCACTTCTGCGCATGGAGTGCATCCACGAGTGCAGGACAGGAGAGTGT 2020
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2108 AAAGGCATTTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAAAATCCGCT 2167
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2019 CGAGTCTTCGTGAGCTATTGTGGCGTCTGCTGCACCGGAGCATTCGAATATGCCCT 1960
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2168 TGCCTATAAATTCAGTTGGNACCCAGCTGTCACCTCCCGTCAGGAAAAATCCCTGCACT 2227
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1959 AAGATATAAGTTCTTGTGTCACCCAGGGGAGTACTTCAACACACTTTCGCTGCAAA 1900
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2228 CTACAAATTTCTTGGTGAGACGATCCATGTAGATGTCATATACTTGTATGAATCAGCAA 2287
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1899 ATATCTGAGTCAGGACAAATTTGCGAGATTTCTCGAGGTG---GGAACTCATGTCAAG 1843
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2288 GAGGCTCAGACTACGAGAGCTTCCAGCTTTTGTCTGGACACTTCCCAAAATCGGAATTC 2347
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1842 TCCTCGCAGCTTGGATTTCCTGCCAGGATTCGCCCTGGAGGGTTCCTCCCAACAGGATTC 1783

CC from (3). The chimeric genes can be used to produce plant cells
CC and seeds with reduced LKR activity, especially in Arabidopsis,
CC corn, soybean, rapeseed, wheat and rice (claimed). LKR is
CC important in controlling free lysine accumulation in plant seeds.
CC LKR activity reduction may be achieved by cloning the claimed
CC fragment, preparing a chimeric gene for cosuppression of LKR,
CC expression of antisense RNA for LKR, and transforming plants with
CC the chimeric gene.
XX
SQ Sequence 323 BP; 104 A; 57 C; 86 G; 72 T; 4 other;

Query Match 3.2%; Score 105.2; DB 20; Length 323;
Best Local Similarity 65.1%; Pred. No. 1.7e-19;
Matches 155; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 2844 AAGCACAACGCGCTACTGCGGAGTTCGGGAAGTTGMAATGCGGAGTCCACCATGCC 2903
DB 11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
3 aagcacactgcgactcttgggaattcggggacatcaagaatggacaacaacaccgct 62

QY 2904 ATGGCGCTGACCGTCTGCGGATTCAGCAATAGGGGCCCTGCTATTGCTAAAGAATAAG 2963
DB 11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
63 atggccaagactgttgggacccctgcagccattggagctctgttaattgaagacaag 122

QY 2964 GTCCAGACGAAGAGAGTGATCAGGCCCTCTGCAACCGGAATCTACGTTCAGCATTTGGAG 3023
DB 11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
123 atcaagacaagaggagtcttaaggcctctcgaagcagagaggtgtatttgcagctttggat 182

QY 3024 ATCTTGAGTCGTCGGGCATCAGCTGTGAGAAAGTGGAGACTTCAAAAGTTCCCTG 3081
DB 11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
183 atattgcaagcatatggtataaagctgagtggagaagcagaatgatcaagaactctg 240

RESULT 13
AAV99566
ID AAV99566 standard; cDNA to mRNA; 308 BP.
XX AC AAV99566;
XX
DT 29-MAR-1999 (first entry)
XX
DE Rice lysine ketoglutarate reductase 3' cDNA.
XX
KW Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW transgenic plant; seed; rice; ss.
XX
OS Oryza sativa.
XX
FH Key Location/Qualifiers
FT CDS 1..129
FT /*tag= a
XX
PN WO9842831-A2.
XX
PD 01-OCT-1998.
XX
PF 27-MAR-1998; 98WO-US06051.
XX
PR 27-MAR-1997; 97US-0824627.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Epelbaum SU, Falco SC, McDevitt RE;
XX
DR WPI; 1999-045139/04.
DR P-PSDB; AAW87764.
XX
XX Nucleic acids and chimeric genes for increasing seed lysine content
PT - comprise sequence encoding all or part of lysine ketoglutarate
PT reductase, useful to improve nutritional quality of seeds from
PT transformed plants
XX
PS Claim 3; Page 197; 231pp; English.

XX This is the nucleotide sequence of a partial (3') cDNA encoding a
CC rice lysine ketoglutarate reductase (LKR) or saccharopine
CC dehydrogenase (SDH) polypeptide (see AAW87764). It was isolated
CC from an expressed sequence tag database by comparison to an
CC Arabidopsis LKR/SDH protein (see AAW87759). Isolated nucleic
CC acid fragments comprising sequences encoding all or part of plant
CC LKRs are new. Also claimed are: (1) a chimeric gene comprising the
CC fragment (or a subfragment) operably linked to a seed specific
CC regulator, where the chimeric gene reduces LKR activity in plant
CC seeds transformed with it; (2) plant cells and seeds in which LKR
CC activity is reduced due to a mutation in the gene encoding LKR or
CC transformation with the chimeric gene; (3) a nucleic acid fragment
CC comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC synthase (DHDDS) substantially insensitive to lysine inhibition is
CC operably linked to a plant chloroplast transit sequence and to a
CC plant seed-specific regulatory sequence; (4) plants comprising in
CC genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC from (3). The chimeric genes can be used to produce plant cells
CC and seeds with reduced LKR activity, especially in Arabidopsis,
CC corn, soybean, rapeseed, wheat and rice (claimed). LKR is
CC important in controlling free lysine accumulation in plant seeds.
CC LKR activity reduction may be achieved by cloning the claimed
CC fragment, preparing a chimeric gene for cosuppression of LKR,
CC expression of antisense RNA for LKR, and transforming plants with
CC the chimeric gene.
XX
SQ Sequence 308 BP; 96 A; 55 C; 72 G; 85 T; 0 other;

Query Match 3.1%; Score 101; DB 20; Length 308;
Best Local Similarity 85.0%; Pred. No. 2.6e-18;
Matches 113; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 2943 CTGCTATTGCTAAAGTAATAGGTCACAGCAAGAGGAGTGATCAGGCTCTGCAACCGGAA 3002
DB 11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
1 ctgctgttctccagaaacaagatccaaagaagagtgatcagcctctggaacctgaa 60

QY 3003 ATCTACGTTCCAGCATTTGGAGATCTTGGAGATCGTCGGGCATCAAGCTGTTGAGAAAGTG 3062
DB 11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
61 attacattccagcttgagatcttggagtcagctgggtatcaagctgagagagtg 120

QY 3063 GAGACTTGAAAGT 3075
DB 11111 11111 11111 11111 11111 11111 11111 11111 11111 11111
121 gagacctgagaat 133

RESULT 14
AAFI3730
ID AAFI3730 standard; cDNA; 1262 BP.
XX AC AAFI3730;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:6253.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO2000056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.

Query Match 2.9%; Score 95.8; DB 20; Length 8160;
best local similarity 78.2%; Pred. NO. 4.9e-16;

Matches	115;	Conservative	0;	Mismatches	32;	Indels	0;	Gaps	0;
QY	1325	GGTATCTCTCAGTGGGCACCTATTTCATAGTTCCCTTATAATGAAGCTTTGGACATCAT	1384						
Db	4580	ggatctctgagcggacacctatttgataagttctgataaacgaagctcttgatgat	4639						
QY	1385	TGAGACAGCTGGAGGTTTCATTTCACCTGGTTAGATGTGAAGTTGCACAAAGCAGCGGATGA	1444						
Db	4640	cgaagcggctgggtgctcatttcttgcctaaatgtgaactgggagagcgcctgatgc	4699						
QY	1445	TATGTCATACTCAGAGCTTGAAGTAGG	1471						
Db	4700	tgaatcgactcagaacttgaagtaag	4726						

Search completed: June 1, 2002, 02:11:27
 Job time: 6987 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2002, 23:20:20 ; Search time 3963.52 Seconds
(without alignments)
17238.501 Million cell updates/sec

Title: US-09-049-304A-120

Perfect score: 3265

Sequence: 1 ATTGTGCGCGCTTCTGCTA.....TCAATAATGCCGATCAGTT 3265

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.px.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg_hum.*

31: em.htg_inv.*

32: em.htg_other.*

33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description

1	3264.6	100.0	3265	6	BD010453
2	3179.8	97.4	3518	8	AF003551
3	1908	58.4	1908	6	BD010454
4	1242.2	38.0	3373	8	ATU95759
5	1236.4	37.9	3195	6	BD010447
6	1232	37.7	3267	8	ATU90522
7	907	27.8	2582	6	BD010452
8	634.6	19.4	1625	8	AF295389
9	593.2	18.2	620	6	BD010451
10	537.8	16.5	1905	8	AY039906
11	522.6	16.0	720	6	BD010455
12	515	15.8	1449	6	BD010458
13	514.4	15.8	1596	8	ATU90523
14	346.6	10.0	15146	8	AF271636
15	300.6	9.2	619	6	BD010450
16	238.4	7.3	533	8	AF042184
17	220.6	6.8	138504	2	AP004095
18	183.6	5.6	429	6	BD010457
19	165	5.1	3098	9	AF229180
20	165	5.1	3233	9	HS007714
21	163.4	5.0	3369	10	BC005420
22	161.8	5.0	3300	10	MMA224761
23	155.8	4.8	3266	6	AX191629
24	152.8	4.7	274	8	AF233461
25	137.4	4.2	33398	8	SPBC3B8
26	114	3.5	42084	3	CBRG44014
27	110.4	3.4	76619	3	AC004722
28	110.4	3.4	88165	2	AC014754
29	110.4	3.4	166722	3	AC008328
30	110.4	3.4	178863	3	AC093195
31	110.4	3.4	268667	3	AE003618
32	105.2	3.2	323	6	BD010449
33	105.2	3.2	323	6	BD010441
34	101.6	3.1	3240	8	AF144424
35	101	3.1	308	6	BD010456
36	99.8	3.1	32218	3	AF038615
37	96.8	3.0	1341	8	SCLYS9
38	96.8	3.0	1594	8	AF191667
39	96.8	3.0	3124	8	SCNR050C
40	95.8	2.9	8160	6	BD010446
41	95.8	2.9	8160	6	BD010446
42	95.8	2.9	111876	8	ATU95758
43	95.8	2.9	199536	8	ATF4110
44	92.8	2.8	372	6	AR014648
45	92.8	2.8	372	6	BD010440

ALIGNMENTS

RESULT 1

BD010453

LOCUS

DEFINITION

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants Chimeric genes and methods for increasing the

lysine content of the seeds of plants.

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

3265 bp DNA linear PAT 31-JAN-2002
Chimeric genes and methods for increasing the lysine content of the seeds of plants Chimeric genes and methods for increasing the lysine content of the seeds of plants.

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

BD010453

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade: Panicoidae; Andropogoneae; Zea.

1 (bases 1 to 3265)

Falco.S.C., Iii.R.E.M. and Epelbaum.S.U.

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Patent: JP 2001502923-A 85 06-MAR-2001;

EI DU PONT DE NEMOURS AND CO

OS Zea mays

PN JP 2001502923-A/85

PD 06-MAR-2001

us-09-049-304a-120.rge

Qy	781	CAGAACACTACACCCCTGGTTTTTTCATGAAGAATGCTCCATATGCATCTGTCACTCGTAA	840
Db	781	CAGAACACTACACCCCTGGTTTTTTCATGAAGAATGCTCCATATGCATCTGTCACTCGTAA	
Qy	841	ACTGTATGTTTGGGAGAAGAGGTTTCCACCATTAATAATATGGATCAGTTTACAGCAAT	900
Db	841	ACTGTATGTTTGGGAGAAGAGGTTTCCACCATTAATAATATGGATCAGTTTACAGCAAT	900
Qy	901	TGATGGAGACTGGTTGTCCCTTTAGTCGGCCCTTTGTGCACATAACTCTGTGATATTGGAGGTT	960

Db 901 TGATGGAGACTGGTGTCCCTTTAGTCGCGCTTGTGACATAACTTGTGATATTGAGAGTT 960

Db 961 CCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCGTTTCTTTTCGTATGATCCTT 102

1021 CTAAGAATTTCATACCATGATGATATGAGAGGTGCCGAGTGGTCTGCTTTGGCTGTGTGACA 108
1091 TTCTCTCCCTACACAGAAATTTCTCTAAAGAGGGCTCCCAAGCATTTTGGAAACATACTACTATCTAGAC 114

1081	Db	TTCTCCCTACAGAAATCTCTAAAGAGGCTCCCAACATTTTGGAAACATACTACTATCTAGAC	11
1141	Qy	TTGTTCCTAGTTTGGCCCTCAGTGAAGCAACCGGAGAACCTTCCTTCTCTACTTGAGAAGAG	120
1141	Db	TTGTTCCTAGTTTGGCCCTCAGTGAAGCAACCGGAGAACCTTCCTTCTCTACTTGAGAAGAG	120
1201	Qy	CTTGCATTGCCACATGCTGGCAGATTAACTCTCTTGTATGAATATATCCTAGGATGAGAA	126
1201	Db	CTTGCATTGCCACATGCTGGCAGATTAACTCTCTTGTATGAATATATCCTAGGATGAGAA	126
1261	Qy	ATATCTATGATAGATTGGCACCGCCCAAAAAACAATPCCATTGCCTGCACAGAAGTATPAGCA	132
1261	Db	ATATCTATGATAGATTGGCACCGCCCAAAAAACAATPCCATTGCCTGCACAGAAGTATPAGCA	132
1321	Qy	CCCTGGTATCTCTCAGTGGGACCTATTGTGATAAGTTCTCTTATAAATGAAGCTTTGGACA	138
1321	Db	CCCTGGTATCTCTCAGTGGGACCTATTGTGATAAGTTCTCTTATAAATGAAGCTTTGGACA	138
1381	Qy	TCATTGAGACAGCTGGAGGTTTCATTCTACCTTGGTTAGATGTGAAGTTGGACAAAGCACGG	144
1381	Db	TCATTGAGACAGCTGGAGGTTTCATTCTACCTTGGTTAGATGTGAAGTTGGACAAAGCACGG	144
1441	Qy	ATGATATGTCTACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTTGGATAAAA	150
1441	Db	ATGATATGTCTACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTTGGATAAAA	150
1501	Qy	TTATTGATTCCTTTGACTTCTTTAGCTTAATGAACATGGTGGAGATCACGATGCCGGGGCAAG	156
1501	Db	TTATTGATTCCTTTGACTTCTTTAGCTTAATGAACATGGTGGAGATCACGATGCCGGGGCAAG	156
1561	Qy	AAATTGAATAGCTCTGAAGATAGGAAAAAGTCAATCAGTATGAAACTGACGTCACAAATTG	162
1561	Db	AAATTGAATAGCTCTGAAGATAGGAAAAAGTCAATCAGTATGAAACTGACGTCACAAATTG	162
1621	Qy	ATAAGGAGGGCCAAAGATTTTAATCTTTGGAGCTGGGAAGAGTCTCTGGCCAGCTGCTG	168
1621	Db	ATAAGGAGGGCCAAAGATTTTAATCTTTGGAGCTGGGAAGAGTCTCTGGCCAGCTGCTG	168
1681	Qy	AGTTTCWGGCATCTTACCAGACATATGTACCTATCTGTTGTATGATACCATTGATGCAGATC	174
1681	Db	AGTTTCWGGCATCTTACCAGACATATGTACCTATCTGTTGTATGATACCATTGATGCAGATC	174
1741	Qy	AAATTCAATGTTATCGTGGCATCTTTGATCAAAAAAGATGCAGAAGAGACAGTTGATGGTA	180
1741	Db	AAATTCAATGTTATCGTGGCATCTTTGATCAAAAAAGATGCAGAAGAGACAGTTGATGGTA	180
1801	Qy	TTGAAAATACAACTGCTACCCAGCTTGATGTGCTGATATTGGAAGCCCTTCAGATCTTGG	186
1801	Db	TTGAAAATACAACTGCTACCCAGCTTGATGTGCTGATATTGGAAGCCCTTCAGATCTTGG	186
1861	Qy	TTTCTCAGGTTGAGGTTTGTAAATAGCTTGTGCTGCTGCTAGTCTTTTCACTGTGCCAATGGCAG	192

|||||
Db 1861 TTCTCAGGTTGAGTTGTAATTAGCTTGCTGCTGCTAGTTTTCATGCTGCCAATGCAG 1920
QY 1921 GAGTATGATGAGTTGGAAGACGACATGGTAACGGCAAGCATGTTGATGAATCCCATGT 1980
Db 1921 GAGTATGATGAGTTGGAAGACGACATGGTAACGGCAAGCATGTTGATGAATCCCATGT 1980
QY 1981 CAAGCTTGAGCCAAAGCTGCCAAAGATGCAGGTGTAACCTATACCTTTGTAATGGGCTAG 2040
Db 1981 CAAGCTTGAGCCAAAGCTGCCAAAGATGCAGGTGTAACCTATACCTTTGTAATGGGCTAG 2040
QY 2041 ATCTGCGATAGATCACTTGATGTCATGAAGATGATGATGAAGCTCATGCGCAAGAGG 2100
Db 2041 ATCTGCGATAGATCACTTGATGTCATGAAGATGATGATGAAGCTCATGCGCAAGAGG 2100
QY 2101 GAAAAATAAAGGCAATTTACATCTTACTGTGCTGATGTCATGTCAGCTGCGCAAGAA 2160
Db 2101 GAAAAATAAAGGCAATTTACATCTTACTGTGCTGATGTCATGTCAGCTGCGCAAGAA 2160
QY 2161 ATCCGCTTGCCCTATAATTCAGTTGGAACCCAGCTGGTGACATCCCGTCCAGGAAAAATC 2220
Db 2161 ATCCGCTTGCCCTATAATTCAGTTGGAACCCAGCTGGTGACATCCCGTCCAGGAAAAATC 2220
QY 2221 CTGAGTCTCAAAATTTCTTGGTGAGACGATCCATGATGATGTCATGTCATGATGAT 2280
Db 2221 CTGAGTCTCAAAATTTCTTGGTGAGACGATCCATGATGATGTCATGATGATGAT 2280
QY 2281 CAGCAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTCTGGAACACTTCCCAAAATC 2340
Db 2281 CAGCAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTCTGGAACACTTCCCAAAATC 2340
QY 2341 GGAATTCCTTGATATATGCTGACCTTTATGTTATCTCCAAAGACATCCACCATATATA 2400
Db 2341 GGAATTCCTTGATATATGCTGACCTTTATGTTATCTCCAAAGACATCCACCATATATA 2400
QY 2401 GGCTACTYTTGTTACGAAGCTTTTGTGATGATGATGATGATGATGATGATGATGAT 2460
Db 2401 GGCTACTYTTGTTACGAAGCTTTTGTGATGATGATGATGATGATGATGATGATGAT 2460
QY 2461 TCTTTGATGCTGCAATCATCTGCTGCAAGATGATGATGATGATGATGATGATGATGAT 2520
Db 2461 TCTTTGATGCTGCAATCATCTGCTGCAAGATGATGATGATGATGATGATGATGATGAT 2520
QY 2521 TCCTTGATGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2580
Db 2521 TCCTTGATGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2580
QY 2581 CTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640
Db 2581 CTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640
QY 2641 AGGAATAGCTGTTAAGACAGCTCAAAACCATCAAGTTCTTGGGACTACATGAAGACTC 2700
Db 2641 AGGAATAGCTGTTAAGACAGCTCAAAACCATCAAGTTCTTGGGACTACATGAAGACTC 2700
QY 2701 AAATACCTAAGGTTGTTGAGGCCATTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 2760
Db 2701 AAATACCTAAGGTTGTTGAGGCCATTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 2760
QY 2761 TGGCTATGGCCACCAATGAGCAGACATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 2820
Db 2761 TGGCTATGGCCACCAATGAGCAGACATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 2820
QY 2821 ACCCGGACGGGCAACCCGCGGAAAGCACCAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2880
Db 2821 ACCCGGACGGGCAACCCGCGGAAAGCACCAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2880
QY 2881 AAATGCGAGGTCACCACTGCCATGGCGTACCGTCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2940
Db 2881 AAATGCGAGGTCACCACTGCCATGGCGTACCGTCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2940
QY 2941 CCCTGCTATTGCTAAAGATAAGGTCAGACGAAAGAGCTGATCAGGCTCTGCAACCGG 3000
|||||

Db 2941 CCCTGCTATTGCTAAAGATAAGGTCAGACGAAAGGAGTGATCAGGCTCTGCAACCGG 3000
QY 3001 AAATCAGCTTCCAGCATTTGGAGATCTTGGAGTCTCGGGCATCAAGCTGGTTGAGAAG 3060
Db 3001 AAATCAGCTTCCAGCATTTGGAGATCTTGGAGTCTCGGGCATCAAGCTGGTTGAGAAG 3060
QY 3061 TGGAGACTTGAAGTTCCCTGATACACAGATAAAGATAGTATGATATAGCAGGCGCATG 3120
Db 3061 TGGAGACTTGAAGTTCCCTGATACACAGATAAAGATAGTATGATATAGCAGGCGCATG 3120
QY 3121 TATCTTTTGTATTAATACTCCGTTCTGGAATATATATTTGTGAACATAAATGTGACAAATAA 3180
Db 3121 TATCTTTTGTATTAATACTCCGTTCTGGAATATATATTTGTGAACATAAATGTGACAAATAA 3180
QY 3181 AAAGAACGGTGGAGTATATTTGTRAGAGCGGCAAGAAACCTCTGATATATGACCTGT 3240
Db 3181 AAAGAACGGTGGAGTATATTTGTRAGAGCGGCAAGAAACCTCTGATATATGACCTGT 3240
QY 3241 CGATATCAAAATAATGCCGATCAGTT 3265
Db 3241 CGATATCAAAATAATGCCGATCAGTT 3265

RESULT 2

AF003551
LOCUS
DEFINITION
Ze mays lysine-ketoglutarate reductase/saccharopine dehydrogenase
bifunctional enzyme mRNA, complete cds.
ACCESSION
AF003551
VERSION
AF003551.2
KEYWORDS
GI:5016094
SOURCE
Ze mays.
ORGANISM
Ze mays.
REFERENCE
1 (bases 1 to 3518)
Papes, F., Kemper, E.L., Cord-Neto, G., Langone, F. and Arruda, P.
Lysine degradation through the saccharopine pathway in mammals:
involvement of both bifunctional and monofunctional
lysine-degrading enzymes in mouse
Biochem. J. 344 Pt 2, 555-563 (1999)
20036337
REFERENCE
2 (bases 1 to 3518)
Cord-Neto, G., Kemper, E.L. and Arruda, P.
Direct Submission
Submitted (09-MAY-1997) Centro de Biologia Molecular e Engenhariaia
Genetica, Universidade Estadual de Campinas, Cidade Universitaria,
Campinas, SP 13083-970, Brazil
3 (bases 1 to 3518)
Cord-Neto, G., Kemper, E.L. and Arruda, P.
Direct Submission
Submitted (27-MAY-1998) Centro de Biologia Molecular e Engenhariaia
Genetica, Universidade Estadual de Campinas, Cidade Universitaria,
Campinas, SP 13083-970, Brazil
Sequence update by submitter
4 (bases 1 to 3518)
Cord-Neto, G., Kemper, E.L. and Arruda, P.
Direct Submission
Submitted (08-JUN-1999) Centro de Biologia Molecular e Engenhariaia
Genetica, Universidade Estadual de Campinas, Cidade Universitaria,
Campinas, SP 13083-970, Brazil
Sequence update by submitter
REMARK
On Jun 8, 1999 this sequence version replaced gi:3157909.
FEATURES
Location/Qualifiers
1..3518
/organism="Ze mays"
/cultivar="AGROCIRES F-352 commercial hybrid"
/db_xref="taxon:4577"
178..3360
/codon_start=1
/product="lysine-ketoglutarate reductase/saccharopine
dehydrogenase bifunctional enzyme"

```

/protein_id="AAC18622.2"
/db_xref="GI:5016095"
/translation="MGSAAETGNDTLLGNGVVGILAEETCNMWERAPLIPSHCARLLL
GGKNGPRVNRIVQPSRRHHDAQYEDAGCEISVDSCGLLIIGIKPKLOMLISD
RAYAFSHTHKAQENPLDKILIEERVSLEFDYELIVDDGDKRSKLAFFAGRAGLID
FLHGGQYSLGYSSTFPLSLGSHMYPSSIAAKAAVIVVAEEIATFGLPSPGICPLV
VFTGVNYSQAGAEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTS
ADMVSHKDPTRQFDKDYAHPEHYTPVHERIAPYASVIVNMYWKEKRPPLLNMDQ
LOQMETSCTPLVGCDDITCDIGSIEFINKSTSLERPFPPDPKSNKSIHDDMEGAGV
CLAVDILPEFSKASQHFNLSRLVASLASVKQPAELPSYLRRAICIAHAGRLTPLY
EYIPMRNTMIDLAPATNPIDPKKISTYLSLGHLPDKELINEADILBTAGSGPHL
VRCEVQSDTMSYSELEVADDTATLDKIDSLSUANEHGGDHDGQIELALKIG
KVNVEYDVTIDKGGPILITILGADVRCRPAEFLASYPDICTYGVDDHDAQHLVIVA
SLYOKDAEETVDGENTATQADVAGISLSDVSQVEVVISLPLFAFHAIAAGVCEI
LKKHMTASYVDESMSNISSOAAKADGVTILCEMGLDPGIDHLSMKMIDBAHARKKI
KAFISYCGGLSPRAANNPALAYKFSWNPAGALRSGKNPVPVYKFGSETIHDVGNLHYS
AKRLRLPEFALEHLNRSLLYGLDYGISKEASTIYRATLYRGEFSEIMVTLSKT
GFFDAANHLQDTSRPTYKGLFDELLNINISTINTDLDIEASGYDDDLIARLLKLC
CKNKEIAYKVTIKFLGLHEETQIPKGSFPFDVICORMEORMAYGHNSQDMVLLHH
EVEVEYPDQPAEKHQATLLFEGKVENGRSTTAMALIVGIPPAIGALLLLKKNKVTQK
VIRLQPELYPALIESSGKIKVEKVT"
BASE COUNT      998 a 735 c 833 g 952 t
ORIGIN

```

```

Query Match      97.4%; Score 3179.8; DB 8; Length 3518;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 3206; Conservative* 1; Mismatches 13; Indels 9; Gaps 1;

QY 1 ATTGTGCGCCGCTTCGTCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTG 60
DB 290 ATTGTGCGCCGCTTCGTCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTG 349
QY 61 TGCAGCCAAAGCACAAAGGAGGATCCATCATGACGCTCAGTATGAGGATGAGGATGCCAGA 120
DB 350 TGCAGCCAAAGCACAAAGGAGGATCCATCATGACGCTCAGTATGAGGATGAGGATGCCAGA 409
QY 121 TTTGAGAGACCTGTGAGATGCGGCTTATCATAGGCATCAACCAACCAACCGCTCGAGA 180
DB 410 TTTGAGAGACCTGTGAGATGCGGCTTATCATAGGCATCAACCAACCAACCGCTCGAGA 469
QY 181 TGATTCCTTCAGATAGACGCTACGCTTCTTTTACACACACACAAAGCCCAAAAGAGA 240
DB 470 TGATTCCTTCAGATAGACGCTACGCTTCTTTTACACACACACAAAGCCCAAAAGAGA 529
QY 241 ATATGCCACTGTAGACAAAGATCCTTGAGAAAGGGTGCTCTTGTGATTATGAGCTAA 300
DB 530 ATATGCCACTGTAGACAAAGATCCTTGAGAAAGGGTGCTCTTGTGATTATGAGCTAA 589
QY 301 TTGTTGGAGATGATGGGAAAAGATCATTAGCATTTGGGAAATTTGCTGGTAGAGCTGGAC 360
DB 590 TTGTTGGAGATGATGGGAAAAGATCATTAGCATTTGGGAAATTTGCTGGTAGAGCTGGAC 649
QY 361 TGATGATTTCTTACATGGCTCGACAGCGATATTGAGCGCTTGGATACCTCGACTCCAT 420
DB 650 TGATGATTTCTTACATGGCTCGACAGCGATATTGAGCGCTTGGATACCTCGACTCCAT 709
QY 421 TTCTCTCTCTGGGAAATCTCATATGATTCCTTCGCTCGCTCGACCCAAAGGCTCGAGTCA 480
DB 710 TTCTCTCTCTGGGAAATCTCATATGATTCCTTCGCTCGCTCGACCCAAAGGCTCGAGTCA 769
QY 481 TTGTCGTTGCAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCGATAGTGT 540
DB 770 TTGTCGTTGCAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCGATAGTGT 829
QY 541 TTGTGTTCACTGGAGTTGGAAGGCTCTCAGGGTGGCAGGAGATATTCAGGTTATTGTC 600
DB 830 TTGTGTTCACTGGAGTTGGAAGGCTCTCAGGGTGGCAGGAGATATTCAGGTTATTGTC 889
QY 601 CCCATACCTTTGTTGATGCTGAGAGCTCCCGGAAATTTTTCAGGCCAGGATCTGCTCA 660
DB 890 CCCATACCTTTGTTGATGCTGAGAGCTCCCGGAAATTTTTCAGGCCAGGATCTGCTCA 949

```

```

QY 661 AGCAATCTCAGTCGACCAAGAGAGATTTTCACTTTATGTTGTTGTTGACCTCTAGAG 720
DB 950 AGCAATCTCAGTCGACCAAGAGAGATTTTCACTTTATGTTGTTGTTGACCTCTAGAG 1009
QY 721 ACATAGTTTCTCACAAGAGTCCCAACAGACAATTTGACAAAGGTGACTATTATGCTCATC 780
DB 1010 ACATAGTTTCTCACAAGAGTCCCAACAGACAATTTGACAAAGGTGACTATTATGCTCATC 1069
QY 781 CAGAACACTACACCCCTGTTTTTTCATGAAGAATTTGCTCCATATGCATCTGTCATCGTAA 840
DB 1070 CAGAACACTACACCCCTGTTTTTTCATGAAGAATTTGCTCCATATGCATCTGTCATCGTAA 1129
QY 841 ACTGTATGTTATGCGAAGAGAGGTTTCCACCAATTAATAATATGATATGATACAGCAAT 900
DB 1130 ACTGTATGTTATGCGAAGAGAGGTTTCCACCAATTAATAATATGATATGATACAGCAAT 1189
QY 901 TGATGGAGACTGTTGCTTTCCTTTAGTCGGGTTTGTGACATAACTTGTGATATGGAGGTT 960
DB 1190 TGATGGAGACTGTTGCTTTCCTTTAGTCGGGTTTGTGACATAACTTGTGATATGGAGGTT 1249
QY 961 CCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCTTTCTTTTCGGTATGATCCTT 1020
DB 1250 CCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCTTTCTTTTCGGTATGATCCTT 1309
QY 1021 CTAGAATTCATACCATGATGATATGGAAGTCCCGAGTGGTCTGCTTGGCTGTTGACA 1080
DB 1310 CTAGAATTCATACCATGATGATATGGAAGTCCCGAGTGGTCTGCTTGGCTGTTGACA 1369
QY 1081 TTCTCCCTACAGAATTCCTTAAAGAGGCTCCCAACAATTTTGGAAACATACTATCTAGAC 1140
DB 1370 TTCTCCCTACAGAATTCCTTAAAGAGGCTCCCAACAATTTTGGAAACATACTATCTAGAC 1429
QY 1141 TTGTTGCTAGTTTGGCTCTAGTGAAGCAACCGGAGAACTTCTTCTTCTTCTTCTTCTTCTTCT 1200
DB 1430 TTGTTGCTAGTTTGGCTCTAGTGAAGCAACCGGAGAACTTCTTCTTCTTCTTCTTCTTCTTCT 1489
QY 1201 CTTGCATTTGCATGCTGCGCAGATTAACTCCTTTTGTATGATATATATATATATATATATATAT 1260
DB 1490 CTTGCATTTGCATGCTGCGCAGATTAACTCCTTTTGTATGATATATATATATATATATATATAT 1549
QY 1261 ATACTATGATAGATTGGCACCCGCAAAACAAATCCATTCCTTCTTCTTCTTCTTCTTCTTCTTCT 1320
DB 1550 ATACTATGATAGATTGGCACCCGCAAAACAAATCCATTCCTTCTTCTTCTTCTTCTTCTTCTTCT 1609
QY 1321 CCCTGCTATCTCTCAGTGGCACCTTATTTGATAGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1380
DB 1610 CCCTGCTATCTCTCAGTGGCACCTTATTTGATAGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1669
QY 1381 TCATTGAGACAGCTGAGGTTTCATTTCATTGTTAGATGTGAAGTTGAAAGTTGGAAGTTGGAAGTT 1440
DB 1670 TCATTGAGACAGCTGAGGTTTCATTTCATTGTTAGATGTGAAGTTGGAAGTTGGAAGTTGGAAGTT 1729
QY 1441 ATGATATGTCATCTCAGAGCTTGAAGTAGAGCAGATGATACCTCCACATTTGGATATAAA 1500
DB 1730 ATGATATGTCATCTCAGAGCTTGAAGTAGAGCAGATGATACCTCCACATTTGGATATAAA 1789
QY 1501 TTATTGATTCCTTGACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1560
DB 1790 TTATTGATTCCTTGACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1849
QY 1561 AAATTGAAATTAGCTCTGAAGATAGGAAAGTCAATGAGTATGAAACTGACGTCACAAATG 1620
DB 1850 AAATTGAAATTAGCTCTGAAGATAGGAAAGTCAATGAGTATGAAACTGACGTCACAAATG 1909
QY 1621 ATAAAGGAGGCCCAAGATTTTAAATTTCTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTG 1680
DB 1910 ATAAAGGAGGCCCAAGATTTTAAATTTCTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTG 1969
QY 1681 AGTTTCTGGGATCTTACCCAGACATATGACCTATGTTGTTGATGACCATGATGACAGATC 1740
DB 1970 AGTTTCTGGGATCTTACCCAGACATATGACCTATGTTGTTGATGACCATGATGACAGATC 2029
QY 1741 AAATTTCATGTTATCGTGGCATCTTTGTTATCAAAAAGATGACAGAGAGAGATGTTGATG 1800

```

Db	2030	AAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGACACAGTTTCATGGTA	2089
Qy	1801	TTGAAATACAACTGCTACCCAGCTTCATGTTGCTGTATATGGAAAGCCTTCAGATCTTG	1860
Db	2090	TTGAAATACAACTGCTACCCAGCTTCATGTTGCTGTATATGGAAAGCCTTCAGATCTTG	2149
Qy	1861	TTTCTCAGGTTGAGGTTGTAATAGCTTGCTGCTGTAGTTTTTCATGCTGCCATTCGAC	1920
Db	2150	TTTCTCAGGTTGAGGTTGTAATAGCTTGCTGCTGTAGTTTTTCATGCTGCCATTCGAC	2209
Qy	1921	GAGTATGCATGAGCTTCAAGAAGCACATGTAACGCCAAGCTATATGTAATGTAATCCATGT	1980
Db	2210	GAGTATGCATGAGTTGAAGAGCACATGTAACGCCAAGCTATGTAATGTAATCCATGT	2269
Qy	1981	CAAACTTGAGCCAAAGCTGCCAAAGATGCAGGTGTAACATATACTTTGTGAAATGGGCTAG	2040
Db	2270	CAAACTTGAGCCAAAGCTGCCAAAGATGCAGGTGTAACATATACTTTGTGAAATGGGCTAG	2329
Qy	2041	ATCCTGGCATAGATCACTTCATGCTCAATGAAGATGATGATGAAGCTCATGCACGAAAG	2100
Db	2330	ATCCTGGCATAGATCACTTCATGCTCAATGAAGATGATGATGAAGCTCATGCACGAAAG	2389
Qy	2101	GA AAAATAAAGGCATTTACATCTTACTGTGGTGGAATGGCATCTCCAGCTGCAGCAAA	2160
Db	2390	GA AAAATAAAGGCATTTACATCTTACTGTGGTGGAATGGCATCTCCAGCTGCAGCAAA	2449
Qy	2161	ATCCGCTTGCCCTATAAATTCAGTTGGAACCCAGCTGGTGCATCTCGGGTCAGGAAAA	2220
Db	2450	ATCCGCTTGCCCTATAAATTCAGTTGGAACCCAGCTGGTGCATCTCGGGTCAGGAAAA	2509
Qy	2221	CTGCACTCACAAATTTCTTGGTGAGACCATCATGTAGATGGTCATACCTTGTATGAT	2280
Db	2510	CTGCACTCACAAATTTCTTGGTGAGACCATCATGTAGATGGTCATACCTTGTATGAT	2569
Qy	2281	CAGCAAGAGCTCGACACTACAGAGCTTCAGCTTTGCTCTGGAACACTTGCCAAATC	2340
Db	2570	CAGCAAGAGCTCGACACTACAGAGCTTCAGCTTTGCTCTGGAACACTTGCCAAATC	2629
Qy	2341	GGAAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGACATCCACCATATATA	2400
Db	2630	GGAAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGACATCCACCATATATA	2689
Qy	2401	GGGCTACTTTCGTTACGAAGGTTTATGAGATATATGGTAACCTTTCCAAAACCTGGGT	2460
Db	2690	GGGCTACTTTCGTTACGAAGGTTTATGAGATATATGGTAACCTTTCCAAAACCTGGGT	2749
Qy	2461	TCTTTGATGCTGCAAAATCATCACTGCTGCAAGATACGTAGTCCCAACATATAGGGTT	2520
Db	2750	TCTTTGATGCTGCAAAATCATCACTGCTGCAAGATACGTAGTCCCAACATATAGGGTT	2809
Qy	2521	TCCITGATGAACACTACTGAATTAATCTCCAAATTAACACGGACTTAGATATTCGAAGCTT	2580
Db	2810	TCCITGATGAACACTACTGAATTAATCTCCAAATTAACACGGACTTAGATATTCGAAGCTT	2869
Qy	2581	CTGCTGGATACGATGATGACCTGATTCGAGACTGTTGAAGCTCGGTTGCAAAAATA	2640
Db	2870	CTGCTGGATACGATGATGACCTGATTCGAGACTGTTGAAGCTCGGTTGCAAAAATA	2929
Qy	2641	AGGAATAGCTGTTAAGACAGTCAAAAACCATCAAGTTCTTTGGGACTACATGAAGAGACTC	2700
Db	2930	AGGAATAGCTGTTAAGACAGTCAAAAACCATCAAGTTCTTTGGGACTACATGAAGAGACTC	2989
Qy	2701	AAATACCTTAAGGGTTGTTCCAGGCCATTTGATGTGATTTGCCAGCGAATGGAACAGAGGA	2760
Db	2990	AAATACCTTAAGGGTTGTTCCAGGCCATTTGATGTGATTTGCCAGCGAATGGAACAGAGGA	3049
Qy	2761	TGGCCTATGGCCAAATGACCAAGACATGCTACTGCTCCACACGAGTCGAGTGAAT	2820
Db	3050	TGGCCTATGGCCAAATGACCAAGACATGCTACTGCTCCACACGAGTCGAGTGAAT	3109
Qy	2821	ACCCGGACGGCAACCCGCCGAAACGACCAACGACGAGCTACTGGAGTTTCGGGAAGGTTG	2880

Db	3110	ACCCGACGGGCAACCCGCCGAAAAAGACCAACGCGACGCTACTGAGTTCTCGGGAAGGTTG	3169
QY	2881	AAATGCGAGTCCACCACTGCCATGCCCTGACCGTCCGCAATTCAGCAGCAATAGGGG	2940
Db	3170	AAATGCGAGTCCACCACTGCCATGCCCTGACCGTCCGCAATTCAGCAGCAATAGGGG	3229
QY	2941	CCCTGCTATTGCTAAAGAATAAGGTCACGACGAAAGAGTGATCAGGCCCTCTGCACCCG	3000
Db	3230	CCCTGCTATTGCTAAAGAATAAGGTCACGACGAAAGAGTGATCAGGCCCTCTGCACCCG	3289
QY	3001	AAATCTACGTTCCAGCATTTGGAGATCTTGGAGTCGCGGCATCAAGCTGGTTGAGAAAG	3060
Db	3290	AAATCTACGTTCCAGCATTTGGAGATCTTGGAGTCGCGGCATCAAGCTGGTTGAGAAAG	3349
QY	3061	TGGAGACTTGAAGTTCCTCGATACACAGATAAAGATAGTATGATATCCAGGGGCACATG	3120
Db	3350	TGGAGACTTGAAGTTCCTCGATACACAGATAAAGATAGTATGATATCCAGGGGCACATG	3409
QY	3121	TATCTTTTGTATTAACCTCGGTTCTCGAATATATTTGTGAACATAAAATGTGAC	3174
Db	3410	TATCTTTTGTATTAACCTCGGTTCTCGAATATATTTGTGAACATAAAATGTGAC	3469
QY	3175	---AAATAAAAGACGGTGAGTATATTGTAAGAGACGGCAAGAAA	3220
Db	3470	AGAAATAAAGACGGAAGTAGTATTTGTAAAAAATAAAAAAAAAA	3518

RESULT 3
BD010454

LOCUS	BD010454	1908 bp	DNA	linear	PAT 31-JAN-2002
DEFINITION	Chimeric genes and methods for increasing the lysine content of the seeds of plants Chimeric genes and methods for increasing the lysine content of the seeds of plants.				

ACCESSION
BD010454

VERSION
BD010454.1

GI:186388277

VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
1 (bases 1 to 1908)

FEATURES

BASE COUNT	537 a	376 c	439 g	556 t
/db_xref="taxon:4577"				

Query Match	58.4%	Score 1908;	DB 6;	Length 1908;
Best Local Similarity	100.0%	Pred No	0.	

QY 1 ATTTGTGCCCGCTTCTGCTAGGAGGAGCAAGAACCGACCTCGAGTAAACCGGATTATTG 60

[illegible]

QY	261.1	GACTGTTGAAGCTCGGGTGTTCGAAAAATAAGAAATAGCTGTTAAGACAGTCAAAACCA	2670
Db	283.5	GAATTATCAGCTTGGACATTCC-----AAGGAGACTGCAGCCAAGCTGCCTCAAAACAA	2888
QY	267.1	TCAAGTTCTTGGGACTACATGAGAGACTCAAAATACCTAAGGTTCTTCGAGCCCATTTG	2730
Db	288.9	TTGTATTCTTGGGTTTCACGAAGAGAGAGGAGGTTCCATCACTGTGTAAAGCGTATTTG	2948
QY	273.1	ATGTGANTTTCGACGCAANTGGAACAGAGAGTGGCCTATGGCCACAATAGCAGACATGG	2790
Db	294.9	ATGCAACTTGTTACCTAATGGAAGAGAACTAGCTTATTCGGGAAATGAACAGGACATGG	3008
QY	279.1	TACTGCTCCACCAAGAGTGGAGTGGAAATACCGGACGGCAACCGCGGAAAAGCAAC	2850
Db	300.9	TGCTTTTGCAATCAGCAAGTAGAAGTGGAAATCCCTTGAAAGCAACAGTATAGAGAAGCAC	3068
QY	285.1	AAGCGACGCTACTGGAGTTCGGGAAGTGTGAAATGCGAGTCCACACACTGCCATCGCGC	2910
Db	306.9	CTGGGACCTTTTGGAAATTCGGGGACATCAAGATGACAAACAACAACCTGTTATGGCCA	3128
QY	291.1	TGACCGTCGGGCAATTCACAGCAATAGGGGCCCTGCTATTGCTPAAAGAATAAGSTCCAGA	2970
Db	312.9	AGACTGTTGGGATCCCTCAGCCCATTTGAGAGCTCGCTGTAAATTTGAACACAGCATCAAGA	3188
QY	297.1	CGAAGGAGTATCATAGGCTCTCGCAACCGGAAATCTAGTTCCTAGCATTTGGAGATCTTGG	3030
Db	318.9	CAAGAGGAGCTTTAAGGCTCTCGAAGCAGAGGTGATTTCGCCAGCTTTGGATATATTGC	3248
QY	303.1	AGTCGTCGGGCATCAAGCTGGTTGAGAAGTGGAGACTTGAAGCTTCCCTG	3081
Db	324.9	AAGCATATGGTATAAAGCTGATGGAAGGAGGAGATGATCAACGAACCTG	3299

RESULT	5
BD010447	
LOCUS	linear DNA 3195 bp PAT 31-JAN-2002
DEFINITION	Chimeric genes and methods for increasing the lysine content of the seeds of plants Chimeric genes and methods for increasing the lysine content of the seeds of plants.
ACCESSION	BD010447
VERSION	BD010447.1 GI:18638820
KEYWORDS	JP 2001502923-A/79.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 3195)
AUTHORS	Falco,S.C., Iii,R.E.M. and Epelbaum,S.U.
TITLE	Chimeric genes and methods for increasing the lysine content of the seeds of plants
JOURNAL	Patent: JP 2001502923-A 79 06-MAR-2001;
COMMENT	EI DU PONT DE NEMOURS AND CO OS Unidentified PN JP 2001502923-A/79 PD 06-MAR-2001 PF 27-MAR-1998 JP 1998543284 PR 27-MAR-1997 US 08/824627 PI SAVERIO CARL FALCO,RAYMOND ERVIN MCDEVITT III, PI SABINE URSULA EPELBAUM PC C12N9/06,C12N9/12,C12N9/88,C12Pl3/08,C12NI5/82 CC
FEATURES	Strandedness: Double; CC Topology: Linear; FH Key Location/Qualifiers FT source 1..3195 FT /organism='Unidentified'. FT Location/Qualifiers 1..3195 /organism='unidentified' /db_xref='taxon:32644'
BASE COUNT	967 a 602 c 750 g 875 t
ORIGIN	

Db	2219	AAATGATCAACGATGCTCATATCAAAAAAGGAAAGTCAAGTCTTTTACCTCTTATTGTG	2278
Qy	2131	GTGGATTGCGATCTCAGCTGAGCAACAATCCGCTTGCTTATATAATTCAGTTGGAAACC	2190
Db	2279	GAGGCTTCCCTCTCTCTGCTGAGCAATAATCCATTAGCATATAAATTTAGCTGGAACC	2338
Qy	2191	CAGCTGGTGCACCTCCGGTCAGGGAATAATCCTGCAGCTACAAAATTTCTTGGTGAGACGA	2250
Db	2339	CTGCTGGAGCAATTCGAGCTGGTCAAAACCCGCCCAATACAAAAGCAACGGCGACATAA	2398
Qy	2251	TCCATGTAGATGTCTAATCTGTATGATCAGCAAGAGGCTCAGACTACGAGAGCTTC	2310
Db	2399	TACATGTGATGGGAAGAATCTCTATGATTCGGCGCAAGATTCGAGTACCTATATCTTC	2458
Qy	2311	CAGCTTTGCTCTGGAACTTCGCCAAATCGGAATTCCTTGATATATATGATGGTACCTTTATG	2370
Db	2459	CAGCTTTGCTGAGTGTGTTTCCAAATCGTACTCTCTGTTTACGGGGAACATATG	2518
Qy	2371	GTATCTCCAAAGAGCATCCACCATATATAGGCTACTTCTGTTACGAAGGTTTTAGTG	2430
Db	2519	GCATCGAGAGCGAAGCAACACGATATTTCTGTTGGAACACTCAGATATGAAGGTTTAGTA	2578
Qy	2431	AGATTATGTAACCTTTTCCAAACTGGGTCTTTGATGCTGCAAAATCATCCACTGCTGC	2490
Db	2579	TGATAATGGCAACACTTTCGAACTTGGATCTTTGACAGTGAAGCAAAATCAAGTACTCT	2638
Qy	2491	AAGTACTAGTCTGCCAACATATAAGGGTTTCTTGTGATGACTACTGAATAATATCTCCA	2550
Db	2639	CCACTGGAAGAGGATTAACGTTTGGTCTCTTTTAAAGTAACTATTAATGAAGATGCGAG	2698
Qy	2551	CAATTAACAGGCTTAGATATTGAACCTTCTGTGTGATACGATGATGACTGATTCGCA	2610
Db	2699	ACAATGAATC-----AGAGCCCTAGCGGGAAGAGAGATGAAGCAAGA	2743
Qy	2611	GACTGTTGAAGCTCGGGTGTTCGCAAAATTAAGAAATAGCTGTGTAAGACAGTCAAAACCA	2670
Db	2744	GAATATCAAGCTTGGACATTC-----AAGGAGACTCGAGCCAAAGCTGCCAAACAA	2797
Qy	2671	TCAAGTTCTTGGGACTACATCAAGAGACTCAAAATACCTAAGGTTGTTCCAGGCCAATTTG	2730
Db	2798	TGTATTTCTTGGGTTCAACGAGAGAGGAGGTTCCTATCACTGTGTAAGCGTATTG	2857
Qy	2731	ATGTGATTGGCAGCGAATGAACAGAGGATGGCTATGGCCCAATATGACAGACATGG	2790
Db	2858	ATGCAACTGTTTACCTAATGAAGAGAACTAGCTTATTTCGGAATTAAGACAGACATGG	2917
Qy	2791	TACTGCTCACACAGAGTCCGAGTGGGAATACCCGAGCGGCAACCCGCCGAAAGACCC	2850
Db	2918	TGCTTTTGCAATCAGAGTAGAAGTGGAAATCTCTTGAAGCAACGTTATAGAGAAGACA	2977
Qy	2851	AAGCGAGCTACTGAGTTCGGGAAGGTTGAAATGGCAGGTCCACCACTGCCATGGCGC	2910
Db	2978	CTGCGACTCTTTTGAATTCGGGACATCAAGATGGCAACAACACACCCGCTATGGCCA	3037
Qy	2911	TGACCGTCCGCAATTCAGCAGCAATAGGGCCCTGCTATTGCTAAAGAAATAGGTCGAGA	2970
Db	3038	AGACTGTTGGGATCCCTCGAGCCATTTGAGCTCTCTGTTTAAATGCAAGCAAGATCAAGA	3097
Qy	2971	CGAAAGAGTGTAGCGCTCTGCAACCGGAAATCTAGCTTCCAGCATTTGGAGATCTTGG	3030
Db	3098	CAAGAGAGTCTTAAGCCCTCTCGAAGCAGAGGTGTAATTTGCCACTTTGGATATATTG	3157
Qy	3031	AGTCTGGCGGCATCAAGCTGGTTGAGAAAGTGA	3064
Db	3158	AAGCATATGTTATAAAGCTCATGGAGAGGCGAGA	3191
RESULT	6		
ATU90522			
LOCUS		3267 bp	mRNA linear
DEFINITION		Arabidopsis thaliana lysine-ketoglutarate reductase/saccharopine	
DEFINITION		dehydrogenase (LKR/SDH) mRNA, complete cds.	
ACCESSION		U90522	

VERSION U90522.1 GI:2076883

KEYWORDS thale cress. thaliana

SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 3267)

AUTHORS Tang, G., Miron, D., Zhu-Shimoni, J.X. and Galili, G.

TITLE Regulation of lysine catabolism through lysine-ketoglutarate reductase and saccharopine dehydrogenase in Arabidopsis

JOURNAL Plant Cell 9 (8), 1305-1316 (1997)

MEDLINE 97432144

PUBMED 9286108

REFERENCE 2 (bases 1 to 3267)

AUTHORS Tang, G., Miron, D., Zhu-Shimoni, J.X. and Galili, G.

TITLE Direct Submission

JOURNAL Submitted (24-FEB-1997) Dept. of Plant Genetics, The Weizmann Institute of Science, Rehovot 76100, Israel

FEATURES

Location/Qualifiers

1..3267

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/tissue_type="seedling hypocotyls"

1..3267

/gene="LKR/SDH"

63..3257

/gene="LKR/SDH"

/EC_number="1.5.1.8"

/EC_number="1.5.1.9"

/function="catalyses the first two steps in lysine catabolism"

/note="bifunctional enzyme; aminoadipic semialdehyde synthase"

/codon_start=1

/product="lysine-ketoglutarate reductase/saccharopine dehydrogenase"

/protein_id="AAB53975.1"

/db_xref="GI:2076884"

/translation="MNSNGHEEEKLGVVGLAETVKNKWRRTPLTPSHCARLLHG GKDRGISRIWVQSAKRHHDALEDEYDCEIDSLSDCGILIGKQPELEMILPERA YAFSTHKAQKNMPLDLKILSERVLCDTELIVDGHGKRLFAFGKAGAGLVDFL HGLQVYLSGYSTFPLSLGAYMPSLSAAKAAVSVGEEIASQGLPLGICPLVVF TGTGNVSLGAQEIFKLLPHTPEVPSKLPFLVKDKIGSNGISTKRVYVIGCIISQ DMVEHKDPSKFDKADYVAHPHYNPFHEKISPYTSVLVNCMTWKSFFPCLJSTKOL QDLTKGLPLVIGCDITCDIGSIEFVNRAFLIDSPFRFNPSNNSYIDMDGDGVLC MAVDILPTERAKESQHEGDIILSGFVGLASMTFISDLPAHLKRACTSYRGELTSLV YIPRRKSNPEEAQNIITANGVSSQRTFNILVSLGHLFDKFLNEALDMEAGGSF HLAKCELGSADAESYLEVGDADKRVLDQIIDSLELAPNEDYISPHREANKISL KIGVQOENEIKERPEMTKSGVLLGAGRVCRPAADFASVRTISSQOHWKYFGAD SEEDTDVHVIVASLYLKDAKETVESISDVEARVLDSDSESLKLYISQDVVLSLLFA SCHAVVAKTCIELKHLVTASYDDETSMLHENAKSAGITILGEMGLDPPGIDPMMAK MTNDAPITKGKVKSTSYCGGLPSRAAANPLAYKFSWNPAGIRAGONPAKYKNSGD IIVDGNKLYDSAAFRVNPAPALECFPNRDSLVYGEHYGIESEATTIFRGTLRYE GFSMLMATLSKLGFDESEANQVLSLTKRITFCALLSNILNKDADNESEPLAGEEISK RLITLGHSTAKAAAKATIVELGFNFEEREVPSLCKSVFDTACTYLMEKLAYSNGEOM VLLHHEVEFEFLSKRIKHTATLLEFGDIRKNGQTTTAMAKTVGIPPAIGALLIEDK IKTGRVLRPLEAEVYLPALDILOQYIKLMEKAE"

BASE COUNT 993 a 620 c 761 g 893 t

ORIGIN

Query Match 37.7%; Score 1232; DB 8; Length 3267;

Best Local Similarity 64.4%; Pred. No. 1.3e-306;

Matches 2000; Conservative 1; Mismatches 1036; Indels 67; Gaps 8;

Qy 15 CTGCTAGAGAGCGAAGACGACCTCGAGTAAACGGGATTTATTCGAGCGCAACACA 74

Db 183 CTTTACACCGTGGGAAGACAGACCGGATTCGCCGATTCGCTTCCAGCATCTGCT 242

Qy 75 AGGAGGATCCATCATGACGCTCAGTATGAGGATGCGAGATTCGAGACCTG 134

Db 243 AAGCGTATCCATCATGATGCGCTTGTGATGAAGATGTTGGGTGTGAAATTTCTGATGATTTG 302
QY 135 TCAGAAATGCGGCTTATCATAGGATCAAAACAAAGCTGCAGATGATTTCTTTTCAGAT 194
Db 303 TCTGATTTGGGCTTATACATTTGGAATCAAAACAAAGCTGCAGTGAATGATTTCTCCAGAG 362
QY 195 AGAGCGTACGCTTTCTTTTTCACACACACAAAGCCCAAAAGAGATATGCCACTGTGA 254
Db 363 AGACATACGCTTTCTTTTTCACATACATCAAGCAGACAGAAAGAACATGCGCTTTGTG 422
QY 255 GACAAGATCCTTGAAGAAAGGTGCTCTTTGATATGAGCTAAATTTGTGAGATGAT 314
Db 423 GATAAAATCTTTCTGAGAGAGTCTTTGTGTGATATGAGCTCATTTGTGGGATCAT 482
QY 315 GGGAAAGATCTACATAGCATTTGGGAAATTTGCTGTAGAGCTGGACATGATAGATTTCTT 374
Db 483 GGGAAAGATTTGGGTTTGGTAAATATGCAGGAGAGCTGGTCTTTGTGACTTCTTGA 542
QY 375 CATGGTCTCGGACAGCGATATTTGAGCCTTGGATCTCGACTCCATTTCTCTCTGGGA 434
Db 543 CACGGACTTGGACAGCGATATCTAAGTCTAGGATCTCAACACCTTTCTCTCGCTCGT 602
QY 435 CAATCTCATATGATCTTCCGCTCGCTGCAGCAAGGCTGCAGTCAATTTGTCTGTCAGAA 494
Db 603 GCATCTGATATGATTTCTCTCATTTGGCTGTGCAAAAGCCGCTGATATTTCTCTGGTGA 662
QY 495 GAGATAGCAACATTTTGGACTTCCATCCGGAATTTTGCAGATAGTGTCTGTTCTACTGA 554
Db 663 GAAATTTGCAAGCCAGGACTGCCATTTAGGATCTGCGCTTCTGATTTCTCTCACACTTTGT 782
QY 555 GTTGGAAAGCTCTCTCAGGCTGCGAGGAGATATTCAGTATTTGCGCCCATACCTTTGTT 614
Db 723 ACAGGAAATGTTTCTCTGGGGCGCAAGAAATTTCAAGCTTCTCTCACACTTTGT 782
QY 615 GATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTCTCAAGCAATCTCAG --- 671
Db 783 GRACCAAGCAACTTCTGACTATTTGTAAAGACAAGGAATTAGTCAAAATGGGAT 842
QY 672 TCGACCAAGAGAGATTTTCAACTTTATGGTGTGTGTGACCTCTAGACATAGTTCT 731
Db 843 TCAACAAAGCGAGTCTATCAAGTATATGTTGATATTAACCAAGCAAGATGTTGAA 902
QY 732 CACAAGATCCACACAGACAATTTCAAAAGGTGACTATATGCTCATCCAGACACTAC 791
Db 903 CACAAGATCCATCAAGTCTATCACAAGGCGGACTATATGACACACCGCAACATAC 962
QY 792 ACCCTGTTTTTCATGAAGAAATTCCTCCATATGATCTGTCATCGTAAACTGTATGAT 851
Db 963 AATCCAGTTTTCCACGAAAGATATCGCCATATACGCTGTCTGTTGTAACCTGTATGAT 1022
QY 852 TGGGAGAAGAGTTTCCACCATTTACTAAATATGGATCAGTTACAGCAATTTGATGAGACT 911
Db 1023 TGGGAGAAGAGTTTTCCCTGTCTTCTGAGCAAAACAGCTTCAAGATTTAAACAAAAA 1082
QY 912 GGTTCCTCTTATGCGGCTTTGTGACATACTTCTGATATTTGAGGTTTCCATTTGAATTT 971
Db 1083 GGACTCCCACTAGTAGGCAATATGATATAACTTGTGACATCGGTGGCTCCATTTGAATTT 1142
QY 972 ATCAACAAGATACATCAATAGAGAGGCTTTCTTTCGGTATGATCTTCTPAAGAATTTCA 1031
Db 1143 GTTAACCGAGCTACTTTAATCGATTTCCCTTTCTTCAGGTTTAAATCCCTCGAACAATTTCA 1202
QY 1032 TACCATGATGATATGGAAGGTTCGCGAGTGTCTGCTGGCTGTGATCTCCCTACA 1091
Db 1203 TACTACGATGACATGGATGGGATGGGCTACTATGATGCTGTGTGATATTTACCCACA 1262
QY 1092 GAATTTCTAAAGAGGCTCCCAACATTTTGGAAACATACTATCTAGACTGTGTGCTAGT 1151
Db 1263 GAAATTTGCAAAAGAGGATCCACGATTTTGGAGATATTTCTTCGGATTTGTGGGTAGT 1322
QY 1152 TTGGCTCAGTGAAGCAACCGGAGAACTTCTCTTCTTACTTGTAGAAGAGCTTGCATTTGCA 1211
Db 1323 CTGGCTTCAATGACTGAAATTTTCAGATCTACACGACATCTGGAAGAGGCTTGCATTAAGC 1382

QY 1212 CATGCTGCAGAGTTAACTCCTTTGTATGAATATATCCCTAGGATGAGAAATCTACTATGATA 1271
Db 1383 TATAGGGGAGAATTTGACATCTTTGTATGAGTATATTCACACGATATGAGGAAGTCAAAATCCA 1442
QY 1272 GATTTGGTACCAGCAAAACAAATCCATTTGCCCTG-----ACAAGAGTATATAGCACC 1322
Db 1443 GAAGAGGCAACAAGATAATATATTCGCCAACGGGTTTCCAGCCAGAGAATCTCAACATA 1502
QY 1323 CTGGTATCTCTCAGTGGGACCTATTTGATAGTTCCCTTTAAATGAAGCTTTGGACATC 1382
Db 1503 TTGGTATCTCTCAGCGGACACCTATTTGATAGTTTCTGATAAAGCAAGCTCTTTGATATG 1562
QY 1383 ATTGAGACAGCTGGAGGTTCAITTTTCACTTTGGTTAGATCTGAAGTTGGACAAGCACGAT 1442
Db 1563 ATCGAAGCGGCTGGTGGCTCAITTTGGCTAAATCTGAACCTGGGCGACAGCGCTGAT 1622
QY 1443 GATATGTCATCTCAGAGCTTTGAAGTAGGAGCAGATGATCTGCCACATTTGGATAAAAT 1502
Db 1623 GCTGAATCGTACTCAGAACTTTGAAGTTGGTGGGATGATAAGAGAGTATTTGGATCAAAATC 1682
QY 1503 ATTGATTCCTTTGACTTCTTTTAGCTAAATGAACATGTTGGAGATCAGGATGCC---GGGCAA 1559
Db 1683 ATTGATTCATTAACCTCGGTTAGCTAATCCAAATGAAGATTTATATATCCCAACATAGAGAA 1742
QY 1560 GAAATGAAATAGTCTGGAAGATAGGAAAGTCAATGAGT-----ATGAAACTGACG 1611
Db 1743 GCAAAATAGATCTCCTCAGGATTTGAAAGTCCAGCAAGAAATGAGATAAAGAGAGAG 1802
QY 1612 TCACAATGATAAAGGAGGCGCAAGATTTTAAATCTTTGAGCTGGAAGAGTCTGTCCGC 1671
Db 1803 CCTGAAATGAGCAAAAAA--TCAGGTGTTTGTATTTCTTGTGCTGGACGTGTGTGTCGCC 1860
QY 1672 CAGCTGCTGAGTTTCTGGCATCTTACCCAGACATATGATAC----- 1711
Db 1861 CAGCTGCTGATTTCTCAGCTTACGTTAGAACCATTTCTGTCACAGCAATGGTACAAACAT 1920
QY 1712 -CTATGTTGTGATGACCATGATGCAGATCAAAATCAATGTTATGTTGGTGGCATCTTTGTATC 1770
Db 1921 ATTTCCGAGCAGACTCTGGAAGAAAAACAGATGTTCAATGTTGTCGCTCTCTGTATC 1980
QY 1771 AAAAGATGAGAGAGACAGATTTGATGGTATTTGAAAAATCAACTGCTTACCCAGCTTGTATG 1830
Db 1981 TTAAGGATGCCAAGAGACGCTTGAAGTATTTTCAGATGTAGAAGCAGTTCCGCTAGATG 2040
QY 1831 TTGCTGATATTTGAAGCCTTTTCAGATCTTCTTCTCAGGTTGAGGTTGTAATTTAGCTTGC 1890
Db 2041 TATCTGATAGTGAAGTCTCCTTAAGTATCTTTCTCAGGTTGATGTTGCTTAAGTTAT 2100
QY 1891 TGCTGCTAGTCTTTCATGCTGCCATTTCCAGGAGTATGCCATAGATGTAAGAACACATGG 1950
Db 2101 TACCTGCAAGTTGTCTGCTGTTGTAGCAAGACATGCAATTTGAGCTGGAAGACATCTCG 2160
QY 1951 TAACGGCAAGCTATGTTGATGAATCCATGTCAAACTTTGAGCCCAAGCTGCCAAGATGCGAG 2010
Db 2161 TCACTGCTAGCTATGTTGATGATGAACGTTCCATGTTACATGAGAAGCTTAAGAGTGCCTG 2220
QY 2011 GTGTAATCTATCTTTGGAATGGGCTAGATCTTGGCATAGATCACTTTGATGTCAATGA 2070
Db 2221 GGATAACGATTTCTAGCGGAAATGGGACTGGACCTCGAATCGATCCCATGATGGCGAAGA 2280
QY 2071 AGATGATTTGATGAAGCTCATGCAAGAAAGGAAATAAAGCATTTACATCTTACTGTG 2130
Db 2281 AATGATCAAGATGCTCTCTATCACAAAAGGGAAGTGAAGTCTTTTACTCTTTATTTG 2340
QY 2131 GTGGATGGCATCTCCAGCTGCAGCAAAACATCCGCTTGCCTTATAAATTTCACTTTGGAACC 2190
Db 2341 GAGGCTTCCCTCTCCTGCTGCAGCAAAATAATCCATTTAGCATATAAATTTACTTGAACC 2400
QY 2191 CAGCTGCTGACTCCGCTCAGGGAAGAAATCTTGCAGCTTACAAATTTCTTTGGTGAAGCA 2250
Db 2401 CTGCTGGAGCAATTCGAGCTGGTCAAAAGCCGCCAAATACAAAAGCAACCGCGACATAA 2460


```

EI DU PONT DE NEMOURS AND CO
OS Unidentified
PN JP 2001502923-A/83
PD 06-MAR-2001
PF 27-MAR-1998 JP 1998543284
PR 27-MAR-1997 US 08/824627
PI SAVERIO CARL FALCO, RAYMOND ERVIN MCDEVITT III, PI SABINE
URSULA BEPBAUM
PC C12N9/06,C12N9/12,C12N9/88,C12P13/08,C12N15/82 CC
Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..620
FT /organism="Unidentified".
FT /organism="Unidentified".
FT Location/Qualifiers
FEATURES
    source
        1..620
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT 162 a 128 c 141 g 189 t
ORIGIN

Query Match 18.2%; Score 593.2; DB 6; Length 620;
Best Local Similarity 99.2%; Pred. No. 6.7e-142;
Matches 617; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 237 GAGATATGCCACTCTTACAGACATCCTTGAAGAAAGGCTGCTCTCTTTGATTATGAG 296
Dy 1 GAGATATGCCACTCTTACAGACATCCTTGAAGAAAGGCTGCTCTCTTTGATTATGAG 60

Qy 297 CTAATTGTTGGAGATGATGGAAAGATCACTAGCATTTGGGAATTTGCTGGTAGAGCT 356
Dy 61 CTAATTGTTGGAGATGATGGAAAGATCACTAGCATTTGGGAATTTGCTGGTAGAGCT 120

Qy 357 GGACTGATAGATTTCTTACATGGTCTCGACAGCAGATATTGAGCCTTGGATCTGACT 416
Dy 121 GGACTGATAGATTTCTTACATGGTCTCGACAGCAGATATTGAGCCTTGGATCTGACT 180

Qy 417 CCATTTCTCTCTGGGCAATCTCATATGATATCTCTCGCTCGACGCAAGCTGCA 476
Dy 181 CCATTTCTCTCTGGGAC-ACTCATATGT-TCCTTCGCTCGCTCGACGCAAGCTGCA 238

Qy 477 GTCATTTGCTTGCAGAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCGCGATA 536
Dy 239 GTCATTTGCTTGCAGAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCGCGATA 298

Qy 537 GTGTTTGTTCACCTGGAGTTGGAAACGCTCTCTCAGGCTGGCCAGAGATATTCAAGTTA 596
Dy 299 GTGTTTGTTCACCTGGAGTTGGAAACGCTCTCTCAGGCTGGCCAGAGATATTCAAGTTA 358

Qy 597 TTGCCCCATACCTTTGTTGATGCTCAGAAAGCTTCCGGAATTTTTCAGGCCAGGAATCTG 656
Dy 359 TTGCCCCATACCTTTGTTGATGCTCAGAAAGCTTCCGGAATTTTTCAGGCCAGGAATCTG 418

Qy 657 TCTAAGCAATCTCAGTCGACCAAGAGATATTTCACATTTTATGTTGTTGTGACCTCT 716
Dy 419 TCTAAGCAATCTCAGTCGACCAAGAGATATTTCACATTTTATGTTGTTGTGACCTCT 478

Qy 717 AGAGACATAGTTTCTCACAAGATCCCAAGAGATTTGACAAAGGTGACTATTATGCT 776
Dy 479 AGAGACATAGTTTCTCACAAGATCCCAAGAGATTTGACAAAGGTGACTATTATGCT 538

Qy 777 CATCCAGAACACTACACCCCTGTTTTTCATGAAGAAATTTGCTCCATATGCTCTGTCATC 836
Dy 539 CATCCAGAACACTACACCCCTGTTTTTCATGAAGAAATTTGCTCCATATGCTCTGTCATC 598

Qy 837 GTAAACTGTATGTTATTTGGAGA 858
Dy 599 GTAAACTGCATGTTATTTGGAAA 620

RESULT 10
AY039906
```

```

LOCUS
DEFINITION Arabidopsis thaliana AT4g33150/F4i10_80 mRNA linear PLN 24-JUN-2001
ACCESSION AY039906
VERSION AY039906.1 GI:14532563
KEYWORDS FLI_CDNA.
SOURCE
ORGANISM Arabidopsis thaliana
Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 1905)
Author: Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Banh,J.,
Bowser,L., Carninci,P., Dale,J.M., Gibson,H.A., Goldsmith,A.D.,
Hayashizaki,Y., Ishida,J., Jiang,P.X., Jones,T., Kamiya,A.,
Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X.,
Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J.,
Pham,P.K., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A.,
Tang,C.C., Toriumi,M., Yamada,K., Yu,G., Yu,S., Shinozaki,K.,
Davis,R.W., Theologis,A. and Ecker,J.R.
Arabidopsis cDNA clones
Unpublished
2 (bases 1 to 1905)
Author: Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Banh,J.,
Bowser,L., Carninci,P., Dale,J.M., Gibson,H.A., Goldsmith,A.D.,
Hayashizaki,Y., Ishida,J., Jiang,P.X., Jones,T., Kamiya,A.,
Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X.,
Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J.,
Pham,P.K., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A.,
Tang,C.C., Toriumi,M., Yamada,K., Yu,G., Yu,S., Shinozaki,K.,
Davis,R.W., Theologis,A. and Ecker,J.R.
Direct Submission
Submitted (07-JUN-2001) Salk Institute Genomic Analysis Laboratory
(SIGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.
The Salk, Stanford, PGEC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Cheuk,R., Chen,H.,
Kim,C.J., Koesema,E., Meyers,M.C., Shinn,P., Banh,J., Bowser,L.,
Dale,J.M., Gibson,H.A., Goldsmith,A.D., Jiang,P.X., Jones,T.,
Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K., Quach,H.L.,
Southwick,A., Tang,C.C., Toriumi,M., Yamada,K., Yu,G., Yu,S.,
Davis,R.W., Theologis,A. and Ecker,J.R.
Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
contributed equally to this work as PIs.
Location/Qualifiers
    1..1905
        /organism="Arabidopsis thaliana"
        /db_xref="taxon:3702"
        /chromosome="4"
        /clone="RAFL05-02-O17(R12911)"
        /note="ecotype: Columbia"
    1..164
        /note="lysine-ketoglutarate reductase/saccharopline"
        /codon_start=1
        /product="AT4g33150/F4i10_80"
        /protein_id="AAK64010.1"
        /db_xref="GI:14532564"
        /translation="MTKSGVLIILGAGRCVPAADFASVRTISSQQWTKYTVGADSE
        EKTQVHVIVASLYLKDAKETVEGSDVEAVLDVSDSESLKYSQVDVLSLLPASC
        HAVAKTCIELKHLVYASYVDDETSMLHEKAKSGITILGMDLPDIDHMAKMI
        NDAHLKGVKVSFTSYCGGLFSPAANNPLAYKFSWNPAIGATRAGQNPAYKNSGDII
        HVDGNLYDLSAARFVNPALPALECFPNRDSLYVGESEATTFRTGLRYEGF
        SWIMATLSKLGFFDSEANQVLTGKRTITGFTGALLSLNLKNDNESEPLAGEEISKRI
```

FEATURES
source5'UTR
CDS

IKLHSHKTAARAAKATIVPLGNEERVPVSLCKSVFADATCYLMEKCLAYSGNEQDWL
LHVEVEFLESKRKEKHDTLLEFGDKNGOTTAMAKTVGIPRAIGALLIEDKIK
TRGVLRPLUEAEVYILFALDILQAYIKLMEKAE
1614..1905
3' UTR
BASE COUNT 634 a 328 c 449 g 494 t
ORIGIN

Query Match 16.5%; Score 537.8; DB 8; Length 1905;
Best Local Similarity 60.5%; Pred. No. 1.5e-127;
Matches 997; Conservative 1; Mismatches 598; Indels 51; Gaps 5;
QY 1465 AAGTAGGACAGATGACTGCCACATTTGGATAAAATATTGATCTCTTTCAGTCTTTAG 1524
DB 1 AGGTTGGTGGGATGATAAGAGAGTAGTATGGATCAAAATCATTTGATTCATTAACGCGTTAG 60
QY 1525 CTAATGAACATGGTGAGATCAGGATGCGG---GGCAAGAAATTTGAATTAGTCTCTGAAGA 1581
DB 61 CTAATCCAATGAAGATTATATATCCCATAGAGAAGCAATAAGATCTCACTGAAGA 120
QY 1582 TAGGAAAAGTCAATGAGTATGA-----AAGTACGCTCAAAATTTGATAAAGGAGGGCCAA 1635
DB 121 TTGGTAAAGTCCAGCAAGAAAATGAGATAAAGAGAAGCTGAAATGACGAAAATAATCAG 180
QY 1636 AGATTTTAAATCTTGAGCTGGAAGAGTCTGCGGCCAGCTGCTGAGTTTCTGGCATCTT 1695
DB 181 GTGTTTGTGATTTGGTGTGAGCGTGTGTGCGCCAGCTGCTGATTTTCTAGCTTCAG 240
QY 1696 ACCAGACATATGTACTAT-----GGTGTGATGACCATGATG 1734
DB 241 TTGAACCAATTCGTACAGCAATGTTACAAAACATATGTCGAGCAGACTCTGAAGAGA 300
QY 1735 CAGATCAATATCATGTTATCGTGGCATCTTTGTATCAAAAAGATGACAGAAGACAGTTG 1794
DB 301 AAACAGATGTTGATGATGATGTCGCTCTGTATCTTAAGGATGCGCAAGAGAGCGGTTG 360
QY 1795 ATGGTATTTGAAAATCAACTGCTACCCAGCTGATGCTGCTGATATTTGAAGCCTTTTCAG 1854
DB 361 AAGGATTTTCAGATGATGAAGCAGTTGCGCTAGATGATCTGATGATGAAGTCTCCTTA 420
QY 1855 ATCTGTTTCTCAGGTTGAGTTGTAATAGCTTGCTGCTGCTAGTTTTCATGCTGCCA 1914
DB 421 AGTATGTTTCTCAGGTTGATGTTGCTTAAGTTTATTACCTGCAAGTTGCTGCTGTTG 480
QY 1915 TTGCAGGAGTATGCATAGAGTTGAAGAAGACATGTTACGCGAAGCTATGTTGATGAAT 1974
DB 481 TAGCAAGACATGCAATGAGCTGAAGAAGCATCTGCTCACTAGCTAGTATGTTGATGATG 540
QY 1975 CCATGTCAAACTTGAGCCCAAGCTGCCAAGATGCGAGGTGTAACATATACCTTTGTAAGTGS 2034
DB 541 AAACGTCCATGTTACATGAGAAGGCTAAGAGTGTGGGATACGATCTPAGCGAAATGG 600
QY 2035 GCCTAGATCCTGGCATAGATCACTTGAATGCAATGAAGATGATGATGAAGCTCATGCAAC 2094
DB 601 GACTGGACCTGGAATCGATACATGATGGCGATGAAATGATCAACGATGCTCATATCA 660
QY 2095 GAAAGGAAAAATAAGGATTTACATCTTACTGTGGTGATGTCATGCACTCCAGCTGCA 2154
DB 661 AAAAGGGAAGTGAAGTCTTTTACCTCTTATTTGAGGAGGCTTCCCTCTCTGCTGTCAG 720
QY 2155 CAACAACATCCGCTTGCTATAAATTCAGTTGGAACCCAGCTGGTGACCTCCGGTCAAGGA 2214
DB 721 CAATAATCCATAGCATATAAATTTAGTGTGAACCTGCTGGAGCAATTCGAGCTGGTC 780
QY 2215 AAAATCCTGAGCTCTACAAATTTCTTGGTGAAGCATCCATGATGATGATCAACTTGT 2274
DB 781 AAAACCCGCCAAATACAAAGCAACGGCGACATAATACATGTTGATGGAAGATCTCT 840
QY 2275 ATGAATCAGAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGTCTGGACACTTGC 2334
DB 841 ATGATTCGCGCAAGATTCAGAGTACCTAATCTTCCAGCTTTTGCATTTGGAGTGTTC 900
QY 2335 CAAATCGGAATTCCTTGTATATATGTTGACCTTTATGGTATCTCCAAAGAGCATCCACCA 2394

DB 901 CAAATCGTACTCTTGGTTTACGGGGAACATTATGGCATCGAGAGCAAGCAACACGA 960
QY 2395 TATATAGGGCTACTYTTTCGTACGAAGGTTTATGTCAGATTTATGTTAAACCTTTCCAAA 2454
DB 961 TATTTGTTGGAACACTCAGATATGAAGGGTTAGTATGATATGATGCGCAACACTTTCCAA 1020
QY 2455 CTGGGTTCTTGTGATGCTGCAAAATCCTCCTGCTGCAAGATAGTCTGCTCCAAACATATA 2514
DB 1021 TTGGATTTCTTGTGACATGAAGCAAAATCAAGTACTCTCCACTGCAAGAGGATACGTTG 1080
QY 2515 AGGTTTCTTGTGAACTACTGAATAAATATCTCCACAATTAACACGGACTTAGATATTG 2574
DB 1081 GTGCTCTTTTAAGTAACATCTAAATTAAGGATGCGACAAATGAATC----- 1126
QY 2575 AAGCTTCTGTTGATACGATGATGACTGATTGCCAGACTGTTGAAGCTCGGGTGTGCA 2634
DB 1127 -AGAGCCCTAGCGGAGAGAGAGATAGCAAGAGAAATATCAAGCTTGGACATTC- 1184
QY 2635 AAAATAAGGAATAGCTGTTAAGACAGTCAAAACCAATCAAGTTCTTTGGGACTACATGAAG 2694
DB 1185 -----AAGGAGACTGCGAGCAAAAGCTGCCAAAACAATTTGATTTCTTTGGGTTCAACGAAG 1239
QY 2695 AGACTCAAAATACCTAAGGGTTGTTCCAGCCCATTTGATGATTTGCCAGCGAATCGAAG 2754
DB 1240 AGAGGAGGTTCCATCACTGTTGTAAGGCGTATTTGATGCAACTTGTACCTAATGGAAG 1299
QY 2755 AGAGGATGCGCTATGCGCAACATGAGCAAGACATGTTGCTCCACCAAGATCGAGG 2814
DB 1300 AGAACTACCTTATTCGGAATGACAGACATGCTGTTGATCAGAGTAGAAG 1359
QY 2815 TGAATACCCGAGCGGCAACCCCGGAAAGACCAAGAGCTTACTTGGAGTTCCGGA 2874
DB 1360 TGAATTCCTTGAAGCAAAACGTATAGAGAAGCACACTCGGACTCTTTTGAATTCGGG 1419
QY 2875 AGTTGAAAATGSCAGGTCACCACTGCCATGCGCTGACCGTCCGCTTCCAGCAGCA 2934
DB 1420 ACATCAAGAAATGACAAACAAACACCGCTATGCGCAAGACTGTTGGGATCCCTGCAAGCA 1479
QY 2935 TAGGGCCCTGCTATTTGCTAAAGAAATAGTCCAGACGAAAGAGTATGATCAGGCCCTCTGC 2994
DB 1480 TTGGAGCTCTGCTGTTAATTTGAAGACAAGATCAAGACAAGAGAGTCTTTAAGSCCTCTCG 1539
QY 2995 AACCGGAATCTACGTTCCAGCATTTGGAGATCTTGGAGTCTGCGGCATCAGCTGGTTG 3054
DB 1540 AAGCAGAGGTTGATTTGCCAGCTTTGGATATATTGCAAGCATATGTTATAAAGCTGATGG 1599
QY 3055 AGAAATGAGAGACTTGAAAGTTCCCTG 3081
DB 1600 AGAAGCAGAAATGATCAAGAAACTCTG 1626

RESULT 11

BD010455

LOCUS

DEFINITION

Chimeric genes and methods for increasing the lysine content of the seeds of plants Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

Chimeric genes and methods for increasing the lysine content of the seeds of plants

[illegible]

RESULT 12

QY	1918	CAGGAGTATGATAGAGTTGAAGAAGCACATGGTAAACGCGAAGCTATGTTGATGAATCCA	1977
Db	371	CAAGACATGCATTTGAGCTGAAGAAGCATCTCGTCACTCTGCTAGCTATGTTGATGATGAA	430
QY	1978	TGTCAAACTTGAGCGAAGCTGCCAAGATGTCAGGTGTAACCTATCTTTGTGAAATGGGCC	2037
Db	431	CGTCCTGTTACATGGAACGCTTAAGAGTGTGGGATAACGATTTAGCGGAAATGGGAC	490
QY	2038	TAGATCTCGCATAGATCACTTTGATGCTCAATGAAGATGATTGATCAAGCTCATCGACAA	2097
Db	491	TGGACCTTGGAAATCGATCCCATGATGCGAAGAAATGATCAAGCATGCTCCTATCAACA	550
QY	2098	AGGGAATAATTAAGGCATTTACATCTTACTGTGTGGATGCCATCTCCAGCTCGACGAA	2157
Db	551	AAGGAAAGTGAAGTCTTTTACCTCTTATTTGTGGAGGGCTTCCCTCTCGTCTCGACAA	610
QY	2158	ACAATCCGCTTGGCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAA	2217
Db	611	ATAATCCATTAGCATATAAATTTAGCTGGAACCCCTCTGGAGCAATTCGAGCTGGTCAAA	670
QY	2218	ATCCTGCAGCTTACAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATG	2277
Db	671	ACCCGCGCAAAATACAAGAACCAACGGCGACATAATACATGTTGATGGGAAGAATCTCTATG	730
QY	2278	AATCAGCAAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGGCTCTGGAACACTTGGCAA	2337
Db	731	ATTCCGCGCAAGATTCGAGTACCTATCTTCCAGCTTTTGGCATTTGGAGTTTTCCTCAA	790
QY	2338	ATCGGAATTCCTTGATATATGTGTACCTTTATGTGTATCTCCAAAGAAGCATCCACCATAT	2397
Db	791	ATCGTGACTCCTTGGTTTACGGGGAACATATATGGCATCGAGAGGAAGCAACAGTAT	850
QY	2398	ATAGGGCTACTVTTCTTTACGAAGCTTTTACTGAGATTATGGTAACCTTTCCAAACTTG	2457
Db	851	TTCGTGGACACTCAGATATGAAGGGTTTAGTATGATAATGGCAACACTTTCCAAACTTG	910
QY	2458	GGTCTTTTGATGCTGCAATCATCCACTGCTGCAAGATACTAGTGTGCTCCAAACATATAAGS	2517
Db	911	GATTCCTTTGACAGTGAAGCAAAATCAAGTACTCTCCACTGGAAGAGGATTACGTTGGTG	970
QY	2518	GTTTCTCTTGCACTACTGATATATCTCCCAATTAACAGCGACTTAGATATTGAAG	2577
Db	971	CTCTTTAAGTAACAATCTTAATTAAGGATGAGACAAATGAATC-----AG	1015
QY	2578	CTTCTGGTGGATACGATGATGACCTGATTTGCCAGACTGTTGAAGCTCGGTGTGCAAA	2637
Db	1016	AGCCCTTAGCGGGGAAGAAGAGATAGAAGAGAAATTAACASCTTGGACATTC-----	1071
QY	2638	ATAAGGAATAGCTTTAAGACAGTCAAAACCATCAAGTTCTTGGGACTACATGAAGAGA	2697
Db	1072	--AAGGAGACTGCAGCCAAAGCTGCCAAAACAATGTATTTCTTGGGGTTCAACGAAGAGA	1129
QY	2698	CTCAATACCTAAGGGTTGTTTCGAGGCCAATTTGATGTGATTTGCCAGCAAGTGGACAGA	2757
Db	1130	GGGAGGTTCCATCACTGTGTAAAGCGTATTTGATGCACTGTTACCTTAATTTGGAAGAGA	1189
QY	2758	GGATGGCCTATGGCCACATAGCAAGACATGTTACTGTCTCCACCACGAAGTCGAGGTGG	2817
Db	1190	AACCTAGCTTTATTCGCGAAATTAACAGAGACATGGTGTCTTTGTCATCAGCAAGTAGAAGTGG	1249
QY	2818	AATACCGGAGCGGCAACCCCGCGAAAGCACCAAGCGACTACTGGATTCGCGGAAGG	2877
Db	1250	AATTCCTTGAAGCAACGTTATAGAGACGCACTCGCACTCTTTGGAATTCGCGGACA	1309
QY	2878	TTGAATATGGCAGGTCCACCCTGCCATGGCGCTGACCGCTGGCATTTCCAGCAGCAATAG	2937
Db	1310	TCAAGATGGCAAAACAACAACCGCTATGGCCAAAGACTCTTGGGATCCCTGCAGCCATTG	1369
QY	2938	GGGCGCTGCTATTCTTAAGAATAAGGTCAGACGAAAGGAGTGTATAGGCGCTCTGCAAC	2997
Db	1370	GAGCTCTGCTTTAATTGAAGACACGATCAAGACAGAGAGTCTTAAGCGCTCTCGAAG	1429

QY	2998	CGGAATCTACGTTCCAGCATTTGGAGATCTTTGGAGTCGCGGCATCAAGCTGGTTGAGA	3057
Db	1430	CAGAGGTTGATATTTCCAGCTTTGGATATATTGCAAGCATATGTTAAAGCTTGATGGAGA	1489
QY	3058	AAGTGGAGACTTGAAGCTTCCCTG	3081
Db	1490	AGGCAGATGATCAAGAACTCTG	1513
RESULT	14		
AF271636			
LOCUS		15146 bp	DNA
DEFINITION		lysine ketoglutarate reductase/saccharopine dehydrogenase	linear
ACCESSION		(LKRSDH) gene, complete cds.	PLN 08-OCT-2000
VERSION		AF271636	
KEYWORDS		AF271636.1	GI:10716964
SOURCE		Zea mays.	
ORGANISM		Zea mays	
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.	
AUTHORS		1 (bases 1 to 15146)	
TITLE		Arruda,P., Kemper,E.L., Papes,F. and Leite,A.	
JOURNAL		Regulation of lysine catabolism in higher plants	
MEDLINE		Trends Plant Sci. 5 (8), 324-330 (2000)	
PUBMED		20371120	
REFERENCE		10908876	
AUTHORS		2 (bases 1 to 15146)	
TITLE		Martinez-Moraes,K.C., Kemper,E.L., da Silva,F.R., Vettore,A.L. and Arruda,P.	
JOURNAL		Direct Submission	
FEATURES		Submitted (23-MAY-2000) CBMEG, Unicamp, Campinas, SP 13083-970, Brazil	
Source		Location/Qualifiers	
gene		1..15146	
CDS		/organism="Zea mays"	
		/cultivar="B73"	
		/db_xref="taxon:4577"	
		4027..15146	
		/gene="LKRSDH"	
		join(4027..4047,4150..4413,4518..4604,4697..4826,5080..5253,5874..5953,6202..6318,6671..6746,7396..7567,7875..7971,8425..8588,9379..9433,9543..9683,9870..10190,10350..10470,10674..10739,10861..10978,11073..11202,11458..11537,11868..12026,12117..12369,12561..12669,13060..13221,13331..13400,13486..13541)	
		/note="bifunctional enzyme: LKR/SDH; lysine 2-oxoglutarate reductase/saccharopine dehydrogenase"	
		/codon_start=1	
		/product="lysine ketoglutarate reductase/saccharopine dehydrogenase"	
		/protein_id="AAG21985.1"	
		/db_xref="GI:10716965"	
		/translation="MCSAATEGNDTLGNGVVGIIAETCNMWRRAPLTTPSHCARLLRGGKNGPRNRIIVQSTRIRHDAQIEDAGCEIUSCEGLIGIKQPKLQMLISDRAYAFSHTHKAQKNMPLDLILBEERHVFEDYELVDGDKRSFGKAGRAGLIDFLHGLQSLGSLYFPFLSCQSHWYFLSAAAKAAIVVAETATGFCGSGICPIVFVFTGNGVQAQAEIKFLHPFVDAEKLFEIQRNLKSQSTKRVQLYGCVVTSRMVSHKDTQFDKDYVAHPHTVPFHERIAPYASVIVNCMYWEKRPPLLMDOQLQOIMETGCPPLGVCCDITCGIGSTEFINKSTSPERFVDPKSNHYDDMEGAVVLAADTILPTFEFSKEASQHNILSRIVASLYQPAELSYLRACATIAHAGRUTPLYEYTPRMTNIDLAPAKTNPLFDKISTLVLSLGHLPDKFLINEALDITITAGSPHLKRVCEVQSDTMSYSELEVGADDTATLDKIISLTSLANEHGDDHAGQETELAKITVKNCEYEDTVIDKGGKILILGAGRCVPAEFLASYPDICTYGVDDHADQTHIVA SLKHMVTAETVDCIENTATQLDADIGSLDLYSQVEVVTSLPASFHAIAAGVCIELKHMVTAETVDESMNSQAQADAGVTILCEMGLDPGDIDHLMDEAHARKGIKAFSTYCGGLPSPAPAAHNPALVYKFSNPNAGALRSKPNPYKFLGETIHWGHNLTAKRLRLRELPALPHEALHNRLNLIYDGLGSKASTIYRATLRKEFSSEMTVLTSTGFEDANHPILQDTSRPTKGFDELLENNLINTDLDIEASGGYDDLTARLLKLGCKNKEATVKTIVTIFLGHUEGTLQPKCGSPFDVICQRMORMAYGNHQDQVLLHHVEVEYFDQQAQKHAQHLLEFGVNGRSTAMALTVPAAIGALLLLKKNVQTKG	

3' UTR
VIRLOPEIYVPALEILESSGIKIVKVT"
13542..15146
/gene="LKRSDH"
BASE COUNT 4043 a 3133 c 3298 g 4672 t
ORIGIN

Query Match 10.0%; Score 326.6; DB 8; Length 15146;
Best Local Similarity 96.0%; Pred. No. 5.6e-73;
Matches 335; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
Qy 1439 GGATGATATGCTATCACTCAGACGCTTGAAGTAGCAGCAGATGATACGCCACATGGATAA 1498
Db 9842 GAAACATATGATACACACCTTCTGAGGTAGGAGCAGATGATACGCCACATGGATAA 9901
Qy 1499 AATATTGATTCCTTGACTCTTTAGCTAATGAACATGTTGGAGATCAGCATGCCGGGA 1558
Db 9902 AATATTGATTCCTTGACTCTTTAGCTAATGAACATGTTGGAGATCAGCATGCCGGGA 9961
Qy 1559 AGAAATTTGAATAGCTCGAAGATAGAAAGTCAATGATGATGAACTGACGTCAAT 1618
Db 9962 AGAAATTTGAATAGCTCGAAGATAGAAAGTCAATGATGATGAACTGACGTCAAT 10021
Qy 1619 TGATAAGAGGGCCCAAGATTTTAACTTGGAGCTGGAAGAGTCTGTGGCCAGCTGC 1678
Db 10022 TGATAAGAGGGCCCAAGATTTTAACTTGGAGCTGGAAGAGTCTGTGGCCAGCTGC 10081
Qy 1679 TGAGTTTCTGTCATCTTACCAGACATATGATGATGTTGTTGATGACCATGATGAGA 1738
Db 10082 TGAGTTTCTGTCATCTTACCAGACATATGATGATGTTGTTGATGACCATGATGAGA 10141
Qy 1739 TCAATTCATGTTATGCTGGCATCTTTGATCAAAAAGATGCAGAGAG 1787
Db 10142 TCAATTCATGTTATGCTGGCATCTTTGATCAAAAAGATGCAGAGAG 10190

RESULT 15
BD010450 619 bp DNA linear PAT 31-JAN-2002
LOCUS Chimeric genes and methods for increasing the lysine content of the
DEFINITION seeds of plants Chimeric genes and methods for increasing the
ACCESSION BD010450.1 GI:18638823
VERSION JP 2001502923-A/82.
KEYWORDS seeds of plants
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 619)
AUTHORS Falco,S.C., Iii,R.E.M. and Epelbaum,S.U.
TITLE Chimeric genes and methods for increasing the lysine content of the
JOURNAL seeds of plants
COMMENT Patent: JP 2001502923-A 82 06-MAR-2001;
EI DU PONT DE NEMOURS AND CO
OS Unidentified
PN JP 2001502923-A/82
PD 06-MAR-2001
PF 27-MAR-1998 JP 1998543284
PI 27-MAR-1997 US 08/824627
PI SAVERIO CARL FALCO, RAYMOND ERVIN MCDEVITT III, PI SABINE
URSULA EPELBAUM
PC C12N9/06,C12N9/12,C12N9/88,C12P13/08,C12N15/82 CC
Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..619
FT /organism='unidentified',
Location/Qualifiers
1..619
/organism='unidentified'
/db_xref='taxon:32644'
BASE COUNT 163 a 120 c 141 g 195 t
ORIGIN

Query Match 9.2%; Score 300.6; DB 6; Length 619;
Best Local Similarity 69.8%; Pred. No. 2e-66;
Matches 435; Conservative 0; Mismatches 184; Indels 4; Gaps 2;
Qy 237 GAGAAATATGCTTCTAGCAAGATCCTTGAAGAAAGGGTGTCTTGTGTGATATGAG 296
Db 1 GAARACATGCTTCTTGGTAAAGATCTAGCTCAGAGGGCATCGTTATATGACTATGAA 60
Qy 297 CTAAATTTGGAGATGATGGGAAAAGATCACTAGCATTTGGAAATTTGCTGGTGGAGCT 356
Db 61 TTAATTTGGTGGGACACTTGGGAAAAGGTTACTTGCATTTGGAAAATTCGCTGGTAGGCT 120
Qy 357 GGACTGATAGATTTCTTACATGGTCTCGGACAGCATATTTAGCCCTTGGATACCTGACT 416
Db 121 GGAATGATCGACTTTTCCGGCGGATTAGCAGACGGTTTTTAAAGTCTTGGATATCAACA 180
Qy 417 COATTTCTCTCTGGGACAATCTCATATGATATCTTCTCGCTCGCTGCAGCCAGGCTGCA 476
Db 181 CCTTTCTTGTCACTTGGATCATCTTACATGTACCTTCCCTGGCTGTCTGAAGGCTGCT 240
Qy 477 GTCAATTCGTTGCAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCCGATA 536
Db 241 GTGATTTCTTGGTGGAA-AAATTCGAGCGAGGATTCGCATTTGGGATTTGTCCCTG 299
Qy 537 GTGTTTGTCTTCACTGGAGTTGGAAACGTCTCTCAGGTGGCGAGGAGATATTCAGTTA 596
Db 300 GTTGTATTATTTACTGGTTTCAGGAAATGTCTTCTGTGTCACAGGAGATATTTAAGCTT 359
Qy 597 TTGCCCCATACCTTTTGTGATCTGAGAGCTTCCCGAAATTTTTCAGGCCAGGAACTGTG 656
Db 360 CTTCCTCATACCTTTTGTGATCCATCTAAACTTACCGGACCTA--CATAGACGAGACCA 416
Qy 657 TCTAAGCAATCTCAGTCGACCAAGAGATATTTCAACTTTATGTTGTGTGTGACTCTCT 476
Db 417 GATCAACCAAGCATGCTTCAAAAAGAGTTTCCCAAGTTTATGTTGTGTGACTGCC 476
Qy 717 AGAGACATAGTTTCTCACAAGGATCCCAAGGATTTGACAAAGGTGACTATTTATGCT 776
Db 477 CAAGACATGTTGTAACCCCAAGATCACGTGATGTTTGGACAAAGCAGACTACTATGCA 536
Qy 777 CATCCAGAACACTTACACCCCTGTTTTCATGAAAGAAATTTGCTCCATATGCTATGCTATC 836
Db 537 CATCTTGAGCATTACAAATCCACTTTTCCATGAAAAAATAGCACCATATGCTATCTGTTAT 596
Qy 837 GTAAACTGATGATTTGGGAGAA 859
Db 597 GTCAATTCATGATTTGGGAAAA 619

Search completed: June 1, 2002, 02:04:30
Job time: 9850 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2002, 22:48:59 ; Search time 2448.65 Seconds
(without alignments)
17996.675 Million cell updates/sec

Title: US-09-049-304A-120
Perfect score: 3265
Sequence: 1 ATTGTGCGCCCTTCTGCTA.....TCAAATAATGCCGATCATGT 3265

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	827.6	25.3	851	10	BG836723
2	520.6	15.9	532	9	AW057000
3	492.8	15.1	856	10	BI952486
4	490.4	15.0	525	9	AW681014
5	433.2	13.3	635	10	BF200255
6	402.2	12.3	449	10	BF317962
7	395.8	12.1	605	10	BG906544
8	387	11.9	607	10	BG724449
9	386.6	11.8	687	10	BI955770
10	376	11.5	644	9	AV928505
11	374.8	11.5	606	9	AV926619
12	363	11.1	594	9	AV923326
13	362.4	11.1	575	10	BG607870
14	355.2	10.9	613	10	BE606591
15	348.8	10.7	731	10	BI421952
16	347.2	10.6	444	10	BG048770
17	347.2	10.6	765	10	BM412142

18 339.6 10.4 827 10 BF271904
19 335.6 10.3 527 10 BF259465
20 327.4 10.0 649 9 AI486200
21 321.4 9.8 719 10 BM413549
22 315.2 9.7 596 10 BG049063
23 308.4 9.4 610 9 AI894899
24 299.6 9.2 627 9 AI488742
25 299 9.2 546 10 BE498116
26 295 9.0 685 9 AI054604
27 294.4 9.0 703 10 BF479283
28 291.4 8.9 765 10 BI435118
29 288.8 8.8 505 9 AL506473
30 287.6 8.8 660 10 BG451844
31 286.4 8.8 502 10 BG370012
32 284.8 8.7 642 10 BG890456
33 280.6 8.6 395 10 BE428366
34 280.4 8.6 727 10 BI422689
35 277.4 8.5 521 9 AW681093
36 276.6 8.5 595 10 BM408787
37 274.8 8.4 608 9 AW034280
38 266.4 8.2 471 10 BM100629
39 265.4 8.1 701 10 BM437056
40 261.4 8.0 497 10 BE435871
41 257 7.9 575 9 AI894874
42 256.4 7.9 521 10 BE434479
43 255.4 7.8 575 9 AW932187
44 253.4 7.8 496 9 AW933092
45 240.4 7.4 568 10 BI921398

ALIGNMENTS

RESULT 1
BG836723

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. .851

/organism="Zea mays"

/cultivar="CO430"

/db_xref="taxon:4577"

/clone="Zm08_04d10"

/clone_lib="Zm08_AAFc"

/clone_ear="Zm08_AAFc"

/tissue_type="Developing kernels (silb crossed)"

/dev_stage="10-11 days post-silk emergence"

/note="Vector: Bluescript SK+/XhoI-ECORI; Site_1: ECORI;

Site 2: XhoI; Field-grown maize ears were silk channel-inoculated in the morning (~10 am) with 1 ml of a Fusarium graminearum macroconidial suspension (500,000 spores/ml) and whole ears were collected and immediately frozen in liquid nitrogen 6 days later."

BASE COUNT	250 a	175 c	204 g	213 t	9 others
ORIGIN					
Query Match	25.3%	Score 827.6;	DB 10;	Length 851;	
Best Local Similarity	98.1%;	Pred. No. 4.6e-217;			
Matches	836;	Conservative 10;	Mismatches 5;	Indels 1;	Gaps 1;
QY	2316	TTTGCTCTGGAAACACTTCCCAAAATCGGAATTCCTTGATATATGTTGACCTTTATGTTATC	2375		
Db	1	TTTTCTCTGGAAACTGCAAAATCGGAATTCCTTGATATATGTTGACCTTTATGTTATC	60		
QY	2376	TCCAAGAAGCATCCACCATATATAGGCTACTTTCGTTACGAAGCTTTTATGTAGATT	2435		
Db	61	TCCAAGAAGCATCCACCATATATAGGCTACTTTCGTTACGAAGCTTTTATGTAGATT	120		
QY	2436	ATGCTAACCCCTTTCCAAAACCTGGTCTTTTGATGCTGCAAAATCATCCACTGCTGCAAGAT	2495		
Db	121	ATGCTAACCCCTTTCCAAAACCTGGTCTTTTGATGCTGCAAAATCATCCACTGCTGCAAGAT	180		
QY	2496	ACTAGTCGTCACCAATATATAGGCTTTCCCTTGATGACTACTGAATATATCTCCACAAAT	2555		
Db	181	ACTAGTCGTCACCAATATATAGGCTTTCCCTTGATGACTACTGAATATATCTCCACAAAT	240		
QY	2556	ACACCGGCTTAGATATTGAAGCTTCTGGTGGATACGATGATGCTGATTCGAGACTG	2615		
Db	241	ASACCGGCTTAGATATTGAAGCTTCTGGTGGATACGATGATGCTGATTCGAGACTG	300		
QY	2616	TTGAAGCTCGGCTTTCGAAAATAGGAATAGCTGTGTAAGACAGTCAAAACCATCAAG	2675		
Db	301	TTGAAGCTCGGCTTTCGAAAATAGGAATAGCTGTGTAAGACAGTCAAAACCATCAAG	360		
QY	2676	TTCTTGGGACTACATGAAGAGACTCAAAATACCTAAGGCTTTGTCGAGCCCATTTGATGTG	2735		
Db	361	TTCTTGGGACTACATGAAGAGACTCAAAATACCTAAGGCTTTGTCGAGCCCATTTGATGTG	420		
QY	2736	ATTTGCCAGCGAATGGAACAGGAGTGGCTATGGCCCAATGAGCAAGACATGTTACTG	2795		
Db	421	ATTTGCCAGCGAATGGAACAGGAGTGGCTATGGCCCAATGAGCAAGACATGTTACTG	479		
QY	2796	CTCCACCAAGTTCGAGGTGGAAATACCCGGACGGGCAACCCGCCGAAAGCACCAGCG	2855		
Db	480	CTCCACCAAGTTCGAGGTGGAAATACCCGGACGGGCAACCCGCCGAAAGCACCAGCG	539		
QY	2856	ACGCTACTGAGTTCGGGAAGGTTGAAATGGCAGGTCCACCACTGCCATGGCGCTGACC	2915		
Db	540	ACGCTACTGAGTTCGGGAAGGTTGAAATGGCAGGTCCACCACTGCCATGGCGCTGACC	599		
QY	2916	GTCCGCATTCACGACGAATAGGGGCCCTGCTATTGTCTAAGAAATAGGTTCCAGACGAAA	2975		
Db	600	GTCCGCATTCACGACGAATAGGGGCCCTGCTATTGTCTAAGAAATAGGTTCCAGACGAAA	659		
QY	2976	GGAGTGATCAGGCCCTCTCCAAACCGGAATCTACGTTCCACATGGAGATCTTGGAGTCG	3035		
Db	660	GGAGTGATCAGGCCCTCTCCAAACCGGAATCTACGTTCCACATGGAGATCTTGGAGTCG	719		
QY	3036	TCGGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGAAGTTCCCTGATACACAGATAAAG	3095		
Db	720	TCGGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGAAGTTCCCTGATACACAGATAAAG	779		
QY	3096	ATAGTATGATATAGCAGGCGCATATGATCTTTGTTATTAACCTCGTCTGGAATATAT	3155		
Db	780	ATAGTATGATATAGCAGGCGCATATGATCTTTGTTATTAACCTCGTCTGGAATATAT	839		
QY	3156	TTGTGNACTAAA	3167		
Db	840	TTGTGNACTAAA	851		

RESULT	2	
AW057000		
LOCUS	660007B05.y1 660 - Mixed stages of anther and pollen	
DEFINITION		
ACCESSION	AW057000	
VERSION	AW057000.1 GI:5932639	
KEYWORDS	EST.	
SOURCE	Zea mays.	
ORGANISM	Zea mays	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.	
AUTHORS	1 (bases 1 to 532)	
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University	
JOURNAL	Unpublished (1999)	
COMMENT	Contact: Walbot V Department of Biological Sciences Stanford University 855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 725 8221 Email: walbot@stanford.edu Plate: 660007 row: B column: 05.	
FEATURES	Location/Qualifiers	
source	1..532	
	/organism="Zea mays"	
	/cultivar="Ohio43"	
	/db_xref="taxon:4577"	
	/clone_lib="660 - Mixed stages of anther and pollen shed"	
	/tissue_type="whole premeiotic anthers to pollen shed"	
	/dev_stage="premeiotic anthers to pollen shed"	
	/lab_host="XLOLR"	
	/note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI; Site_2: XhoI; Anther and pollen cDNA library. Directionally sequenced with 5' end at the EcoRI site. Created by Amie Franklin."	
BASE COUNT	157 a 116 c 111 g 148 t	
ORIGIN		
Query Match	15.9%; Score 520.6; DB 9; Length 532;	
Best Local Similarity	99.8%; Pred. No. 2.2e-132;	
Matches	520; Conservative 1; Mismatches 0; Indels 0; Gaps 0;	
QY	2046 GGCATAGTCACTTGATGTCATCAATCAAGATGATTGATGAAGCTCATGCACGAAAGGAAAA	2105
Db	12 GGCATAGTCACTTGATGTCATCAATCAAGATGATTGATGAAGCTCATGCACGAAAGGAAAA	71
QY	2106 ATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAACAATCCG	2165
Db	72 ATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAACAATCCG	131
QY	2166 CTTGCTATAAATTCAGTTGGAACCCAGCTGGTGCATCTCCGCTCAGGGAATAATCTTGA	2225
Db	132 CTTGCTATAAATTCAGTTGGAACCCAGCTGGTGCATCTCCGCTCAGGGAATAATCTTGA	191
QY	2226 GTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGTTGATCTTCTATGATACGCA	2285
Db	192 GTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGTTGATCTTCTATGATACGCA	251
QY	2286 ACAGGCTCAGACTACGAGAGCTTCCAGCTTTTGTCTGGAACACTTCCCAATTCGGAA	2345
Db	252 ACAGGCTCAGACTACGAGAGCTTCCAGCTTTTGTCTGGAACACTTCCCAATTCGGAA	311
QY	2346 TCCTTGATATATGGTGACCTTTATGATATCTCCAAAGAGCATCCACCATATATAGGCT	2405
Db	312 TCCTTGATATATGGTGACCTTTATGATATCTCCAAAGAGCATCCACCATATATAGGCT	371
QY	2406 ACTYTTCTGTTACGAAGCTTTTACTGAGATATATGGTAACCTTTTCCAAAACACTGGTCTT	2465

Db 372 ACTCTCTGCTACGAAGGTTTGTAGTGAGATTATGTTACCCCTTCCAAACCTGGGTTCTTT 431
Qy 2466 GATGCTGCAAAATCATCCACTGCTCAAGACTACTAGTCGTCACATATAAGGGTTTCCTT 2525
Db 432 GATGCTGCAAAATCATCCACTGCTCAAGACTACTAGTCGTCACATATAAGGGTTTCCTT 491
Qy 2526 GATGAACTACTGAATAATATATCTCCACAAATTAACAGGACTT 2566
Db 492 GATGAACTACTGAATAATATATCTCCACAAATTAACAGGACTT 532
RESULT 3
BI952486
LOCUS
DEFINITION
BI952486 856 bp mRNA linear EST 19-OCT-2001
HVSDNA0014 (Blumeria infected) Hordeum vulgare cDNA clone
HVSDNA0006K11f, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
REFERENCE
AUTHORS
1 (bases 1 to 856)
Wing, R., Close, T.J., Klein, H., A., Wise, R., Chin, A., Begum, D.,
Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons
J., Oates, R. and Main, D.
Development of a genetically and physically anchored EST resource
for barley genomics: Blumeria infected Morex (compatible) seedling
cDNA library
Unpublished (2001)
CONTACT: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 435
Seq primer: AATTAACTCTCACTAAAGG
High quality sequence start: 11
High quality sequence stop: 602.
Location/Qualifiers
1. .856
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone_lib="HVSDNA0006K11f"
/clone_lib="Hordeum vulgare green seedling EST library"
HVSDNA0014 (Blumeria infected)
/tissue_type="green seedling leaf"
/lab_host="TJC121"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; Morex (mla) plants were greenhouse grown in the R
Wise lab at Iowa State University, Ames, IA; 7 day old
green seedlings were infected with isolate 5874 of
Blumeria graminis f. sp. hordei, and leaves were harvested
24, 48 and 72 hr post-inoculation and snap frozen (Wise).
In the TJ Close lab at the University of California,
Riverside, total RNA was prepared from each sample pool,
equal quantities of all three RNA pools were combined,
poly(A) RNA was purified from the mixture, one primary
unamplified cDNA library was made, and 1 million pfu were
in vivo excised to give pBluescript SK(-) cDNA phagemids
(Chin). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
The sequence has been trimmed to remove vector sequence

and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Klein H, A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT 237 a 162 c 235 g 222 t
ORIGIN

Query Match 15.1% Score 492.8; DB 10; Length 856;
Best Local Similarity 79.5%; Pred. No. 1.4e-124;
Matches 620; Conservative 1; Mismatches 153; Indels 6; Gaps 3;
Qy 1740 CAAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAAGAGACAGTTCATGCTT 1799
Db 7 CAGGTACATGTTGTTGTGGCATCTCTGTATCAAAAAGATGCAAGAGACAGTTCACGGA 66
Qy 1800 ATTGAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTTGGAAGCCCTTTCAGATCTT 1859
Db 67 ATAAAGATGCAACAGCAGCTCAGCTGATGTTTTCAGATACTGAAAGTCTTTCGAATCTT 126
Qy 1860 GTTCTCAGGTTGAGGTTGTAATAGCTTCTGCTGCTAGTATTTTCATCTGCTCCATTGCA 1919
Db 127 GTTTCAGGTTGATGCTGCTAGTACAGTGGGGGCTGCTAGTATTTTCATCTGCTCCATTGCA 186
Qy 1920 GGAGTATGATAGCTTGAAGAGACATGTTGAAGGCAAGCTATGTTGATGAATCCATG 1979
Db 187 AGAGTATGATAGCTTGAAGAGACATGTTGTCACAGCAAGCTATGTTGATGATTCATG 246
Qy 1980 TCAACTTTGAGGCAAGCTGCCAAAGATGCAAGTGTGATGATGATGATGATGATGATGATG 2039
Db 247 TCAAGTTGGAGCAAGCTGCCAAAGATGCAAGTGTGATGATGATGATGATGATGATGATG 306
Qy 2040 GATCCTGGCATAGATCATTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2099
Db 307 GATCCTGGCATAGATCATTGATGATGATGATGATGATGATGATGATGATGATGATGATG 366
Qy 2100 GGAAATTAAGGCAATTTACATCTTACATCTTACATCTTACATCTTACATCTTACATCTT 2159
Db 367 GGAAATTAAGGCAATTTACATCTTACATCTTACATCTTACATCTTACATCTTACATCTT 426
Qy 2160 AATCCGCTTGGCTATPAAATTCAGTTGGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2219
Db 427 AATCCACTGGCTATPAAATTCAGTTGGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 486
Qy 2220 CTTGAGTCTACAAATTTCTTGGTGGAGCATCCATGATGATGATGATGATGATGATGATG 2279
Db 487 CTTGAGTCTACAAATTTCTTGGTGGAGCATCCATGATGATGATGATGATGATGATGATG 546
Qy 2280 TCAGAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTCTGCTGGAACACTTTCGCAAT 2339
Db 547 TCAGAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTCTGCTGGAACACTTTCGCAAT 606
Qy 2340 CGGAATTCCTTGATATATGTCGACCTTTATGATGATGATGATGATGATGATGATGATG 2399
Db 607 CGAATTCCTTGATATATGTCGACCTTTATGATGATGATGATGATGATGATGATGATG 664
Qy 2400 AGGCTACTTCTTACGAGGTTTATGATGATGATGATGATGATGATGATGATGATGATG 2459
Db 665 AGGCTACTTCTTACGAGGTTTATGATGATGATGATGATGATGATGATGATGATGATG 721
Qy 2460 TTCTTTGATGCTCAATATCATCCATGCTGCAAGATATAGTCGTCACATATATAGGGT 2519
Db 722 GTTTTGTGCTTACATGATCCCTTGGT-CAAGGAAGTAAATCGCCCAATTTATGGGT 780

RESULT 4
AW681014
LOCUS

AW681014 525 bp mRNA linear EST 19-JUL-2000

DEFINITION WSL_8_B05_b1_A002 Water-stressed 1 (WSL) Sorghum bicolor cDNA, mRNA sequence.

ACCESSION AW681014

VERSION AW681014.1 GI:7554719

KEYWORDS EST.

ORGANISM sorghum.

Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 525)
Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt
, L.H.
An EST database from Sorghum: water-stressed plants

TITLE Unpublished (2000)

JOURNAL Contact: Cordonnier-Pratt MM

COMMENT Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.

Seq primer: JEN REV

High quality sequence stop: 517

POLYA-No. Location/Qualifiers

1. 525

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Water-stressed 1 (WSL)"

/notes="Organ: Mix of 5-week old plants on days 7 & 8 after

water was withheld; Vector: Lambda Zap; Site_1: XhoI;

Site_2: EcoRI; The library was made from poly-A RNA in the

cloning vector lambda ZAP II. Clones to be sequenced were

prepared by mass excision."

155 a 102 c 122 g 146 t

BASE COUNT

ORIGIN

Query Match 15.0%; Score 490.4; DB 9; Length 525;

Best Local Similarity 96.9%; Pred. No. 4.9e-124;

Matches 500; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1735 CAGATCAAAATTCATGTTATCGTGGCATCTTTGTATCAAAAGATGCAGAGACAGCTTG 1794

DB 10 CAGATCAAAATTCATGTTATCGTGGCATCTTTGTATCAAAAGATGCAGAGACAGCTTG 69

QY 1795 ATGGTATTCAAAATCAACTGCTACCCAGCTTGATTTGCTGATATTGGAGCCTTTTCAG 1854

DB 70 ATGGTATTCAAAATCAACTGCTACCCAGCTTGATTTGCTGATATTGGAGCCTTTTCAG 129

QY 1855 ATCTTGTCTTCAGGTTGAGGTTGTAATAGCTTCTGCTGCTAGTTTTCATCTGCCA 1914

DB 130 ATCTTGTCTTCAGGTTGAGGTTGTAATAGCTTCTGCTGCTAGTTTTCATCTGCCA 189

QY 1915 TTCCAGGATGTCATAGATTGAAGAAGCATGTTAAGCATGTTAAGCATGTTAAGCATGTTAAGCAT 1974

DB 190 TTCCAGGATGTCATAGATTGAAGAAGCATGTTAAGCATGTTAAGCATGTTAAGCATGTTAAGCAT 249

QY 1975 CCATGTCAAACTTACGCCAGGTCGCAAGATGCGAGTGTACTATCTTTCTGAAATGG 2034

DB 250 CCATGTCAAACTTACGCCAGGTCGCAAGATGCGAGTGTACTATCTTTCTGAAATGG 309

QY 2035 GCCTAGATCTGTCATAGATCACTTGTATCTCAATGAAGATGATTTGATGAAGCTCATGCAC 2094

DB 310 GCCTAGATCTGTCATAGATCACTTGTATCTCAATGAAGATGATTTGATGAAGCTCATGCAC 369

QY 2095 GAAAGGAAAAATAAAGGCAATTTACATCTTACTTGTGGTGGATGGCACTCCAGCTGCAG 2154

DB 370 GAAAGGAAAAATAAAGGCAATTTACATCTTACTTGTGGTGGATGGCACTCCAGCTGCAG 429

QY 2155 CAACAATCCGCTTCCCTATTAATTCAGTTGGAACCCAGCTGGTGCACCTCGGTCAGGGA 2214

DB 430 CAACAACCCACTTCCCTATTAATTCAGTTGGAACCCAGCTGGTGCACCTCGGTCAGGGA 489

QY 2215 AAATCCCTGAGCTTACAAATTTCTTGGTGAGACGA 2250

DB 490 AAATCCCTGAGCTTACAAATTTCTTGGTGAGACGA 525

RESULT 5

BF200255 635 bp mRNA linear EST 16-APR-2001

LOCUS WHE2254_G05_N10ZE Triticum monococcum vegetative apex cDNA library

DEFINITION Triticum monococcum cDNA clone WHE2254_G05_N10, mRNA sequence.

ACCESSION BF200255

VERSION BF200255.1 GI:11114917

KEYWORDS EST.

SOURCE Triticum monococcum.

ORGANISM Triticum monococcum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 635)

Anderson, O.D., Chao, S., Dubcovsky, J., Echenique, V., Han, P.S., Hsia

, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L.,

Stamova, B. and Tong, J.C.

The structure and function of the expressed portion of the wheat

genomes - Vegetative apex cDNA library from Triticum monococcum

Unpublished (2001)

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: oanderson@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: Stratagene T3 primer.

Location/Qualifiers

1. 635

/organism="Triticum monococcum"

/cultivar="DW92"

/db_xref="taxon:4568"

/clone="WHE2254_G05_N10"

/clone_lib="Triticum monococcum vegetative apex cDNA

library"

/tissue_type="vegetative shoot apex"

/dev_stage="Three weeks-old plants"

/lab_host="E. coli XL0LR"

/note="Vector: Lambda pbk-CMV (Lambda Zap Express), tissue,

excised phagemid; Site_1: EcoRI; Site_2: XhoI; The library

total RNA, and poly(A) RNA were prepared, a cDNA library

was made, and the cDNA clones were in vivo excised at the

University of California, Davis (V. Echenique, B. Stamova

, J. Dubcovsky). Plasmid, DNA preparations and DNA

sequencing were performed in the OD Anderson Lab (all

other authors)."

BASE COUNT 197 a 134 c 152 g 152 t

ORIGIN

Query Match 13.3%; Score 433.2; DB 10; Length 635;

Best Local Similarity 82.2%; Pred. No. 3.4e-108;

Matches 511; Conservative 1; Mismatches 104; Indels 6; Gaps 1;

QY 2259 GATCGTCATACCTGATGATGAATCAGCAAGAGCTCAGACTACGAGAGCTTCAGCTTTT 2318

DB 14 GATCGTCATACCTGATGATGAATCAGCAAGAGCTCAGACTACGAGAGCTTCAGCTTTT 73

QY 2319 GCTCTGGAACACTTGCCTGCAATTCCTTATATATGTTGACCTTTATGATCTCC 2378

```

Db 74 GCCCTGGAACACACTTGGCCAAATCGAAATTCCTTGATGATGGAGACCTGTATGGGATCTCC 133
QY 2379 AAAGAAGCATCCACCATATATAGCGCTACTYTTCTGTTAGCAAGGTTTGTAGTGAATATG 2438
Db 134 AAAGAAGCATCTACTGTATAGTCCACTCTTCGTTATGAAGGATTTAGTGAATATG 193
QY 2439 GTAACTCTTCCAAACTGGGTTCTTTGATGCTGCAAAATCATCCACTGCTGCAAGATATC 2498
Db 194 GCTATCTCTGGCGAAATGGGTTTTTGTGCTGCTGAAATCATCCACTGCTCAAGAACT 253
QY 2499 AGTCGTCCTCAACATATAGGTTTCTCTTGATGAACACTGAAT-----AATATCTCCACA 2552
Db 254 AATCGCCCAACATATAGGATTTTCTCAAGCAACTCTTAAATGACAACAATGTATCCACA 313
QY 2553 ATTAAACAGGACTTAGATATTGAAGCTTCTGGTGGATAGCATGATGACTGATTCGCCAGA 2612
Db 314 TCCAAACACAAGTAATAGTGAAGAACTGGAGGACATGATGACTGATTTCAAGA 373
QY 2613 CTGTTCAAGCTCGGGTGTGCAAAATTAAGGAAATAGCTGTTAAGACAGTCAAAACCATC 2672
Db 374 CTCATGATGCTTGGGCATTTGCAAGAAAGGAACTAGCTGCAAGATACTCAAAACCATC 433
QY 2673 AAGTTCTTGGGACTACATGAAGAGACTCAAAATACCTAAGGTTGTTCGAGCCCATTTGAT 2732
Db 434 AAGTTCTTGGGCTGCATGAGGAGACACAGATTCCTAAGGATTTGTTCAAGTGCAATTCAGT 493
QY 2733 GTGATTTGCCCGCAATGAACAGAGGATGGCCTATGCCACAATGAGCAAGACATGGTA 2792
Db 494 GTATTTGCCAAGCAATGAACAGAGATGGCCTATGCCACAATGAGCAAGATGGTA 553
QY 2793 CTGCTCCACACAGAGTGGAGTGAATACCGGACGGGCAACCGCGGAAACACCAA 2852
Db 554 CTGCTCCACCATGAGTGGAGTGGAGTACCCGACGGGCAACCGCGGAGAGACCAA 613
QY 2853 GCGAGCTACTGGAGTTCCGGA 2874
Db 614 GCTAGCTGCTGGAGTTCCGGA 635

RESULT 6
LOCUS BF317962 449 bp mRNA linear EST 21-NOV-2000
DEFINITION OVI_10_A06_b1_A002 Ovary 1 (OVI) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BF317962
VERSION BF317962.1 GI:11266499
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 449)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
J.L.H.
TITLE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 414
POLYA=No.
Location/Qualifiers
1..449
/organism="Sorghum bicolor"
FEATURES
source

```

```

/db_xref="taxon:4558"
/clone_lib="Ovary 1 (OVI)"

```

```

/Note="Organ: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: pBluescript II from Lambda
zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda zap II.
Clones to be sequenced were prepared by mass excision."

```

```

BASE COUNT 136 a 91 c 98 g 124 t
ORIGIN

```

```

Query Match 12.3%; Score 402.2; DB 10; Length 449;
Best Local Similarity 93.3%; Pred. No. 1e-99;
Matches 419; Conservative 1; Mismatches 29; Indels 0; Gaps 0;

```

```

QY 2236 TTCCTGGTGGAGACCATCATGTAGATGCTCATACCTTGTATGAATCAGCAAGAGGCTCA 2295
Db 1 TTCCTGGAGAGACGATCCATGTAGATGCTCATACCTTGTATGAATCAGCAAGAGGCTCA 60
QY 2296 GACTACGAGAGCTTCCAGCTTTTCTCTGGAACACCTTGCACAAATCGGAATTCCTTGTAT 2355
Db 61 GACTACGAGAGCTTCCAGCTTTTCTCTGGAACACCTTGCACAAATCGGAATTCCTTGTAT 120
QY 2356 ATGGTGACCTTTATGGTATCTCCAAAGAACATCCACCATATATAGGCTACTYTTGCTT 2415
Db 121 ATGGAGACCTTTATGGTATCTCCAAAGAACATCTACCATATACAGGCTACTCTTCTGTT 180
QY 2416 ACAAGGTTTGTAGTATGATTAACACGCTTTCACCAACTGGGTTCTTGTATGCTGCA 2475
Db 181 ACAAGGTTTGTAGTATGATTAACACGCTTTCACCAACTGGGTTCTTGTATGCTGCA 240
QY 2476 ATCATCACTGCTGCAAGATAGTCTGCTCAACATATAGGCTTTCCTTGTATGAACACT 2535
Db 241 ATCATCACTGCTGCAAGAACTTAATCGCCCAACATATAGGCTTTCCTTGTATGAACACT 300
QY 2536 TGAATAATATCTCCCAATTAACACGCTTGTAGATTAACACGCTTTCCTTGTATGATG 2595
Db 301 TTAATAATATCTCCCAACTTAACACGCTTGTAGATTAACACGCTTTCCTTGTATGATG 360
QY 2596 ATGACCTGATTCGACAGCTGTTGAGCTCGGTTGTGCAAAATAGGAATAGCTGTTA 2655
Db 361 ATGAAATCATTCGACAGCTGTGCAAGCTCGGTTGTGCAAGATAGGAATAGCTGCTA 420
QY 2656 AGACAGTCAAAACCATCAAGTTCTTGGGA 2684
Db 421 AGACAGTCAAAACCATCAAGTTCTTGGGA 449

```

```

RESULT 7
LOCUS BG906544 605 bp mRNA linear EST 05-JUN-2001
DEFINITION TaLR1150D12R TaLR1 Triticum aestivum cDNA clone TaLR1150D12 5',
BG906544 mRNA sequence.
ACCESSION BG906544
VERSION BG906544.1 GI:14314220
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 605)
AUTHORS Cloutier,S., Dong,G. and Walsh,A.
TITLE Wheat functional genomics- Thatcher Lr1 cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@em.agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).

```

Tue Jun 4 15:23:10 2002

Average inset size is >2.2 kb
Plate: 150 row: D column: 12
Seq primer: M13 Reverse.

FEATURES

source
1. 605
/organism="Triticum aestivum"
/cultivar="Thatcher Lr1"
/db_xref="taxon:4565"
/clone="TaLr1150D12"
/clone_lib="TaLr1"
/tissue_type="Leaf tissue"
/dev_stage="14 Days old"
/lab_host="E. coli XL0R"
/note="Vector: Lambda ZapII; mass excised in plasmid vector pBK-CMV (Stratagene).; Site_1: EcoRI; Site_2: XhoI; mRNA obtained from wheat NIL Thatcher Lr1 24 hours after inoculation with leaf rust pathogen Puccinia triticina race BBB carrying the avirulence gene Avr1."
187 a 135 c 135 g 148 t

BASE COUNT
ORIGIN

Query Match 12.1%; Score 395.8; DB 10; Length 605;
Best Local Similarity 81.1%; Pred. No. 7.1e-98;
Matches 473; Conservative 1; Mismatches 103; Indels 6; Gaps 1;
QY 2131 GTGGATTCATCTCCAGCTGAGCAACAAATCCCTTGGCTATATAATTCAGTTGGACCC 2190
DB 23 GTGGACTTCATCTCCAGCTGAGCAACAAATCCCTTGGCTATATAATTCAGTTGGACCC 82
QY 2191 CAGCTGGTCACTCCGGTCAGGGAATAATCCCTGAGTGTACAAATTTCTTGGTGAGACGA 2250
DB 83 CAGCAGGTGCCATCCGAGCAGGAAGAAATCTGCTGTCTACAAATTTCTTGGAGACCA 142
QY 2251 TCATGTAGATGTCTATACCTTGTATGAATCAGCAAGGCTCAGACTACGAGCTTC 2310
DB 143 TCAATGTGGTGTAGTAAATATACGAGTCAGCAAGGCTCAGACTACGAGCTTC 202
QY 2311 CAGCTTTTGTCTGGAACACTTCCCAATCGGAATTCCTTGTATATATGTGTGACCTTTATG 2370
DB 203 CGCTTTTGTCTGGAACACTTCCCAATCGGAATTCCTTGTATATATGTGTGACCTTTATG 262
QY 2371 GTATCTCCAAAGAGCATCCACCATATATAGGCTACTTTCGTATCAGAAAGTTTATGTG 2430
DB 263 GGATCTCCAAAGAGCATCTACCGTATATAGGTCCTACTCTGTTATGAAGATTTAGTG 322
QY 2431 AGATTATGGTAACCTTTCCAAACTGGTCTTTTGTATCTGCAATCATCCACTGCTGC 2490
DB 323 AGATCATGGCTATCTCTGGCGAAATTTGGGTTTTTGTATGCTGAAATCATCCACTGCTAC 382
QY 2491 AAGATCATGTCTGCCAACATATAAGGTTTCTCTGATGAACCTACTGAAT-----AATA 2544
DB 383 AAGAACTAATGCCCAACATATAGGATTTTCTCAACGAATCTCTAATGTCAACAATG 442
QY 2545 TCTCCAAATTAACACGACTTATGATTAATGAAGCTTCTGTGTGATGATGACCTGA 2604
DB 443 TATCCACATCTAACACAAAGTAAATGGTGAAGAAATCGGAGACATGACATGAAGTCA 502
QY 2605 TTGCCAGACTTCAAGCTCGGTGTTGGCAAAATTAAGAAATAGCTGTTTAAGACAGCTCA 2664
DB 503 TTTCAAGACTCATGATGCTTGGGCAATGCAAGAAAGGAAGTACTAGCTGTCAAGACTCA 562
QY 2665 AAACATCAAGTTCTTGGGACTACATGAAGAGACTCAAAATACC 2707
DB 563 AAACATCAAGTTCTTGGGCTGATCAGGAGACACAGATTC 605

RESULT 8
BG724449/c BG724449 607 bp mRNA linear EST 08-MAY-2001
LOCUS
DEFINITION EST0006 Dactylis leaf DDR1-cDNA Dactylis glomerata cDNA clone
588-T3 similar to Lysine ketoglutarate reductase, mRNA sequence.
ACCESSION BG724449

VERSION
KEYWORDS
SOURCE
ORGANISM

BG724449.1 GI:14007845
EST
Orchard grass.
Dactylis glomerata

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1. (bases 1 to 607)
Trejo Calzada, R. and O'Connell, M.A.
Drought induced transcripts in leaves of Dactylis glomerata
Unpublished (2001)
Contact: Mary A. O'Connell
Department of Agronomy and Horticulture
New Mexico State University
MSC 30, P.O. Box 30003, Las Cruces, NM 88003, USA
Tel: 505 646 5172
Fax: 505 646 6041
Email: moconnel@nmsu.edu
Insert Length: 750 Std Error: 0.00
Seq primer: T3
High quality sequence stop: 607.
Location/Qualifiers
1. 607
/organism="Dactylis glomerata"
/db_xref="taxon:4509"
/clone="588-T3"
/clone_lib="Dactylis leaf DDR1-cDNA"
/dev_stage="drought-stressed"
/note="Organ: Leaf; Vector: pGEM-T"
179 a 134 c 125 g 169 t

FEATURES

source
1. 607
/organism="Dactylis glomerata"
/db_xref="taxon:4509"
/clone="588-T3"
/clone_lib="Dactylis leaf DDR1-cDNA"
/dev_stage="drought-stressed"
/note="Organ: Leaf; Vector: pGEM-T"
179 a 134 c 125 g 169 t

Query Match 11.9%; Score 387; DB 10; Length 607;
Best Local Similarity 78.8%; Pred. No. 1.9e-95;
Matches 475; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 531 CCGATAGTGTGTTGCTCAGTGGAGTTGGAAACGCTCTCTCAGGTCGCGAGGAGATATTC 590
DB 603 CCAATAGTATTGTTGATTCACGGGATCTGGAAATGTTTCGAGGCGCACAGGAGATATTT 544
QY 591 AAGTTATTCGCCCATACCTTTGTTGATGCTGAGAAAGCTTCCCGAAATTTTCAGGCCAGG 650
DB 543 AAGCTATTGCCCCATACCTTCCTGATGCTGACAAACTCTCTGAGATTTCTAAGGGCAAG 484
QY 651 AATCTGTCTAAGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTTATGGTGTGTGTG 710
DB 483 AGTCGGTCT---CCACATGAGTCAACCCAGGAGAGCATTTCCAACTATATGGATGTGTG 427
QY 711 ACCTCTAGACATAGTTTCTCAAGGATCCACAGACAAATTTGACAAAGTGACTAT 770
DB 426 ACATCGAGAGATATGCTCTACCCCAAGGATCCAGCAATTTTCGACAAAGATGACTAT 367
QY 771 TATGCTCATCCAGAACACTACACCCCTGTTTTTATGAAGAATTTGCTCCATATGCACTCT 830
DB 366 TATGCTCACCCAGAACACTACAACTCTGTTTCCATGAAGAGATTACTCCATATGCACTCT 307
QY 831 GTCACTGTAACCTGTATGTTTGGGAGAGAGTTTCCACCAATTTACTAAATATGGATCAG 890
DB 306 GCGATGTTTAATGTATGTACTGGAGGAGGTTTCCACGCTATTGAGCAGTGAACAG 247
QY 891 TTACAGCAATTTGATGGAGACTGTTGCTCTTTAGTCGGGTTTGTGACATATGCTGTGAT 950
DB 246 TTACAACTGCTGACGACAAATGGATGCTCTTTGTTGGCTTTGGGATATACTTTGTGAC 187
QY 951 ATTGAGGTTTCCATGAATTTTATCAACAGACTACATCAATAGAGAGGCTTTCTTCTCGG 1010
DB 186 ATTGAGGTTTCCATGAATTTTGTGAATAGAGTATCATCATGAGAGGCTTTCTTCTCGG 127
QY 1011 TATGATCTCTTAAAGAAATTCATACCATGATGATGAAAGGTCGCCGAGTGCTGCTGTG 1070
DB 126 TACGATCTCTTACTAATTTATATACATGATGATGAAAGGCAATGGTGTGATTTGCTTA 67

ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
1 (bases 1 to 644)
Sato, K., Saisho, D. and Takeda, K.
Barley EST sequencing project in NIG and Okayama Univ
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

REFERENCE
AUTHORS Sato, K., Saisho, D. and Takeda, K.
TITLE Barley EST sequencing project in NIG and Okayama Univ
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source
1. 644
/organism="Hordeum vulgare subsp. vulgare"
/cultivar="Haruna Niho"
/db_xref="taxon:112509"
/clone="basd101"
/clone_lib="K. Sato unpublished cDNA library, cv. Haruna
Niho second leaf stage seedling leaves"
/tissue_type="seedling leaves"
/dev_stage="second leaf stage"
130 a 181 c 141 g 192 t

BASE COUNT
130 a 181 c 141 g 192 t

ORIGIN
11.5%; Score 376; DB 9; Length 644;
Best Local Similarity 81.3%; Pred. No. 2.1e-92;
Matches 436; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Query Match 11.5%; Score 374.8; DB 9; Length 606;
Best Local Similarity 78.3%; Pred. No. 4.5e-92;
Matches 492; Conservative 0; Mismatches 97; Indels 39; Gaps 2;

QY 2536 TCAATAATATCTCCACATTAACAGGACTTAGATATTGAGCTTCTGGTGATACGATG 2595
DB 643 TCAACAATGATFCCACATCTACACAAAGGTAATGTTGTAAGAACTGGAGACATGATG 584
QY 2596 ATGACCTGATTGCCAGCTGTTGAGCTCGGGTGTGCAAAATTAAGAAATAGCTGTTA 2655
DB 583 ATGAAGTATTCAAGACTAATGATGCTTGGGCATTGCAAGAGAAAGAACTAGCTGTC 524
QY 2656 AGACAGTCAAAACCATCAAGTCTTGGGACTACATGAAGAGACTCAATACCTAAGGCTT 2715
DB 523 AGATATCTAAACCACTAAAGTCTTGGGGCTGCATGAGGAGACAGAGATTCCTAAGGAT 464
QY 2716 GTTCAGGCCATTTCGATGTTTCCAGCGAATGGAACAGAGGATGCCCTATGGCCACA 2775
DB 463 GTTCAAGTGCATTCAGTGTATTGCCAACGATGGAACAGAGATGCCCTATGGCCACA 404
QY 2776 ATGAGCAAGACATGTACTGTCTCCACCAGAGTTCGAGTGGATATCCCGACGGGCAAC 2835
DB 403 ATGAGCAGGATATGTACTGTCTCCACCATGAAGTGGAGTACCCGACGGGGCAG 344
QY 2836 CCGCGAAAGCACCAGCGGCTACTGGAGTTCGGAAGTTGAAATGCGAGGTCCA 2895
DB 343 CCACCGAAGAACCCAGCAAGCGCTGCTGGAGTTTCGGAAGACCCAGAACGCGAGTCAA 284
QY 2896 CCACCTGCCATGGCGTGCAGCGCTGGCATTCCAGCAGCAATAGGGGCCCTGTATTGCTAA 2955
DB 283 CCACCGCCATGGCCCTCACCGTTCGGGTACCGGACGATAGGAGCCCTGTCTTGTCTCC 224
QY 2956 AGAATAAGTCCAGAGGAAGAGTGTATCAGGCCTCTGCAACCGGAAATCTACGTTCCAG 3015
DB 223 AGAACAAGTTCAGAGGAAGCGTGTATCCGGCCCTCTGCAACCGGAGATCTACATCCCTG 164
QY 3016 CATTGGAGATCTTGGAGTCTCGGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGA 3071
DB 163 CGCTGGAGATCTTGGATGCTCGGGCATCAAGCTGATCGAGAGAGTGCAGACCTGA 108

RESULT 11
AV926619
LOCUS

DEFINITION
AV926619 K. Sato unpublished cDNA library, cv. Haruna Niho second
leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA
clone basd23b07 5', mRNA sequence.
AV926619 GI:18222416
EST.
Hordeum vulgare subsp. vulgare.
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
1 (bases 1 to 606)
Sato, K., Saisho, D. and Takeda, K.
Barley EST sequencing project in NIG and Okayama Univ
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source
1. 606
/organism="Hordeum vulgare subsp. vulgare"
/cultivar="Haruna Niho"
/db_xref="taxon:112509"
/clone="basd23b07"
/clone_lib="K. Sato unpublished cDNA library, cv. Haruna
Niho second leaf stage seedling leaves"
/tissue_type="seedling leaves"
/dev_stage="second leaf stage"
177 a 117 c 150 g 161 t 1 others

BASE COUNT
177 a 117 c 150 g 161 t 1 others

ORIGIN
11.5%; Score 374.8; DB 9; Length 606;
Best Local Similarity 78.3%; Pred. No. 4.5e-92;
Matches 492; Conservative 0; Mismatches 97; Indels 39; Gaps 2;

QY 1467 GTAGGACAGATGATCTGCCACATTTGGATAAAATTTATGATTCCTTGACCTTTTAGCT 1526
DB 17 GTAGGACAGATGATGACGATFACATTTAGACAAGATCATTTGATTCCTTGACCTTTTAGCT 76
QY 1527 AATGAACATGTTGAGATCAGCATCCCGGCAAGAAATTTGAATTTAGCTCTCAAGATAGGA 1586
DB 77 AATCGCACCTGGAGATCTTAATTCGCG-----CCGAGATATCTCTTAAGATAGGA 127
QY 1587 AAAGTCAATGATGATGAACCTGACGTCACAATTTGATAAAGGAGGCCAAAGATTTTAAT 1646
DB 128 AGATGACCGAATCGGGAATTTGATGACAGCATGATTAAGTAGGCAAAAGTTTAAATC 187
QY 1647 CTTCGAGCTGGAAGAGTCTGTGCGGCGAGCTGCTGAGTTTCTGGCATCTTACCCAGACATA 1706
DB 188 CTTCGAGCTGGAAGAGTGTGTCGCGCGGCTGCTGAGTTTCTTAACATCTTACCAAAACAT- 246
QY 1707 TGTACCTATGTTGATGACCATGATCGACATCAAAATTCATGTTATCGTGGCATCTTTG 1766
DB 247 -----CGACCAAGTACATGTTGTTGTCGCACTCTG 277
QY 1767 TATCAAAAAGATGCGAAGAGACAGATTTGATGGTATTGAAAAATACAACCTGTACCCAGCTT 1826
DB 278 TATCAAAAAGATGCGAAGAGACAGATTTGACGGAATAAAGAAATGCAACAGCAGCTCAGCTC 337
QY 1827 GATGTTGCTGATATTGGAAAGCCTTTCACATCTTGTTCCTCAGGTTGAGGTTGTTAATAGC 1886
DB 338 GATGTTTTCAGACTACTGAAAGTCTTTCGAATCTTGTTCACAGTTTGTGTCGTAGTCAGC 397
QY 1887 TGTGTCGCTGCTAGTTTTCATGCTGCCATTCGAGGATGATGATAGAGTTTGAAGAAGCAC 1946
DB 398 TGTGTCGCTGCTAGTTTTCATGCTGCCATTCGAGGATGATGATAGAGTTTGAAGAAGCAC 457
QY 1947 ATGTTAACGGCAAGCATATGTTGATGAATCCATGTCATAAAGTTCGACCAAGCTGCCAAGAT 2006

```

Db 458 TTGGTCACAGCAGCTATGTTGATGATTCATGTCATGTCGCAAGGCTGGCGCAAGGA 517
QY 2007 GCAGGCTGAACCTACTTCTGAAATGGGCTAGATCCTGGCATAGATCATTGATGTCA 2066
Db 518 GCAGGTGAACCTACTTCTGAAATGGGCTTGTGATCCTGGCATAGATCATTGATGTGCG 577
QY 2067 ATGAAGATGATTGATGAAGCTCATGCAC 2094
Db 578 ATGAAATGATTGATGAAGCACATGCTC 605

RESULT 12
AV923326
LOCUS
DEFINITION
AV923326 K. Sato unpublished cDNA library, cv. Haruna Nijo second
leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA
clone basd12101 5', mRNA sequence.
ACCESSION
AV923326
VERSION
AV923326.1 GI:18219105
KEYWORDS
EST.
SOURCE
Hordeum vulgare subsp. vulgare.
ORGANISM
Hordeum vulgare subsp. vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
REFERENCE
1 (bases 1 to 594)
Sato, K., Saisho, D. and Takeda, K.
Barley EST sequencing project in NIG and Okayama Univ
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
Location/Qualifiers
1..594
/organism="Hordeum vulgare subsp. vulgare"
/cultivar="Haruna Nijo"
/db_xref="taxon:112509"
/clone_lib="basd12101"
/clone="basd12101"
Nijo second leaf stage seedling leaves"
/tissue_type="seedling leaves"
/dev_stage="second leaf stage"
BASE COUNT 173 a 114 c 149 g 157 t 1 others
ORIGIN
1
Query Match 11.1%; Score 363; DB 9; Length 594;
Best Local Similarity 79.8%; Pred. No. 8e-89;
Matches 454; Conservative 0; Mismatches 85; Indels 30; Gaps 1;

QY 1566 GAATTACTCTGAAGATAGGAAAGTCAATGATGATGAACTGACGCTCACAATTGATAA 1625
Db 56 GAGATATCTCTAAAGATAGGAAAGTCAAGGATGCGGAATGATGACACCATGGATAA 115
QY 1626 GGAGGGCCCAAGATTTTAATCTTGGAGCTGGAAGAGTCTGTGGCCAGCTGCTGAGTTT 1685
Db 116 GTAGGACCAAGAGTTTAAATCTTGGAGCTGGAAGAGTTTGTGCGCGGCTGCTGAGTTT 175
QY 1686 CTGGCATCTTACCCAGACATGATCTACCTATGTTGGTGGATGACCATGATGACGATCAAAAT 1745
Db 176 CTAACATCTTACCAACAT-----CGACCAGTA 205
QY 1746 CATGTTATCGTGCATCTTTGTATCAAAAAGATGCAGAAGACAGTTGATGTTATGAA 1805
Db 206 CATGTTGTTGTCATCTCTGTATCAAAAAGATGCAGAAGACAGTTGACGGAATAAG 265
QY 1806 AATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAGCCCTTCAGATCTTGTCTT 1865

```

```

Db 266 AATGCAACAGCAGCTCAGCTCGATGTTTTCAGATACTGAAAGCTTTTCGAATCTTGTTC 325
QY 1866 CAGTTTCAGAGTTGTAATTAGCTTTCCTGCTGCTAGTTTTCATGCTGCCATTGCAAGGAGTA 1925
Db 326 CAGGTTGATGCTGCTAGTACGCTTTCCTGCTGCTAGTTTTCATGCTGCCATTGCAAGGAGTA 385
QY 1926 TGCATAGAGTTGAAGAAGCAGCATGGTAACGGCAAGCTATGTTGATGAATCCATGTCAAAC 1985
Db 386 TGCATAGAGCTCAAGAAGCAGCTTGGTCACAGCAAGCTATGTTGATGATTCATGTCANAG 445
QY 1986 TTGAGCCCAAGCTGCCAAAGATGCAGGTGTAACATATATCTTTGTGAAATGGCCCTAGATCCT 2045
Db 446 TTGAGCAAGCTGCGCAAGCAGCAGGTGTAACATATATCTTGTGAAATGGCCCTAGATCCT 505
QY 2046 GGCATAGATCATTGATGCTCAATGAAGATGATTGATGAAGCTCATGCAAGGAGGAAAA 2105
Db 506 GGCATAGATCATTGATGCTCAATGAAGATGATTGATGAAGCTCATGCAAGGAGGAAAA 565
QY 2106 ATAAAGCATTACATCTTACTTGTGTTGGTGG 2134
Db 566 ATAAAGCGTTTACATCTTCTGTGTTGGTGG 594

RESULT 13
BG607870
LOCUS
DEFINITION
BG607870 575 bp mRNA linear EST 17-APR-2001
WHE2473_E07_J13ZS Triticum monococcum early reproductive apex cDNA
library Triticum monococcum cDNA clone WHE2473_E07_J13, mRNA
sequence.
ACCESSION
BG607870
VERSION
BG607870.1 GI:13657853
KEYWORDS
EST.
SOURCE
Triticum monococcum.
ORGANISM
Triticum monococcum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE
1 (bases 1 to 575)
Anderson, O.D., Chao, S., Dubcovsky, J., Echenique, V., Han, P.S., Hsia
, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L.,
Stamova, B. and Tong, J.C.
The structure and function of the expressed portion of the wheat
genomes - Early reproductive apex cDNA library from Triticum
monococcum
Unpublished (2001)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: StrataGene SK primer.
FEATURES
Location/Qualifiers
1..575
/organism="Triticum monococcum"
/cultivar="DV92"
/db_xref="taxon:4568"
/clone="WHE2473_E07_J13"
/clone_lib="Triticum monococcum early reproductive apex
cDNA library"
/tissue_type="Early reproductive apex"
/dev_stage="Seven week-old plants"
/lab_host="E. coli XL0LR"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; The tissue, total RNA, and
poly(A) RNA were prepared from apex at double-ridge stage
to terminal-spikelet stage during transition from
vegetative state to flower state, a cDNA library was made,
and the cDNA clones were in vivo excised at the
University of California, Davis (V. Echenique, B. Stamova

```


, J. Dubcovsky). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors).

BASE COUNT 173 a 111 c 139 g 152 t
 ORIGIN

Query Match 11.1%; Score 362.4; DB 10; Length 575;
 Best Local Similarity 78.0%; Pred. No. 1.2e-88;
 Matches 479; Conservative 0; Mismatches 96; Indels 39; Gaps 2;

Qy 1469 AGGAGCAGATGATCTGCGCACATTTGGATAAAATTTATGATTCCTTGACTCTTTAGCTAA 1528
 Db 1 AGGAGCAGATGATGACATACATTTAGACAAAGATCATCGATTCCTTGACTCTTTAGCTAA 60

Qy 1529 TGAACATGGTGGAGATGATGCGGCGCAAGAAATGAATAGCTCTGAAGATAGGAAA 1588
 Db 61 TGCACACCGCTGGAGATCTTAATGCC-----ACCGAGATATCTTAAGATAGGAG 111

Qy 1589 AGTCAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1648
 Db 112 AGTCAGCAATGTGGAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 171

Qy 1649 TGGAGCTGGAAGATCTGTCGCGCAGCTGCTGAGTCTTCTGATCTTACCAGACATATG 1708
 Db 172 TGGAGCTGGAAGATCTGTCGCGCAGCTGCTGAGTCTTCTGATCTTACCAGACATATG 228

Qy 1709 TACCTATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1768
 Db 229 -----CGACCAAGTACATCTTGTGTCGATCTCTGTA 261

Qy 1769 TCAAAAGATGCAAGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1828
 Db 262 TCAAAAGATGCAAGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 321

Qy 1829 TCTTCTGATATGGAAGCTTTCAGATCTTGTCTCAGCTTGTGAGTGTGAATAGCTT 1888
 Db 322 TGTTCAGATGATGAAAGCTTTCGAATCTTGTTCACAGTGTGATGATGATGATGATGATG 381

Qy 1889 GCTGCTGCTAGTTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1948
 Db 382 GCTACCTGCCAGTTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 441

Qy 1949 GGTACGCGCAGCTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2008
 Db 442 GGTACGCGCAGCTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 501

Qy 2009 AGGTGTAATCTTCTGTAATGGGCTTAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2068
 Db 502 AGGTGTAATCTTCTGTAATGGGCTTAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 561

Qy 2069 GAAGATGATGATG 2082
 Db 562 GAAATGATGATG 575

RESULT 14
 BE606591
 LOCUS WHE0903_C03_F052s Wheat 5-15 DAP spike cDNA library Triticum
 DEFINITION aestivum cDNA clone WHE0903_C03_F05, mRNA sequence.
 ACCESSION BE606591
 VERSION BE606591.1 GI:9883755
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 1; Triticaceae; Triticum.
 1 (bases 1 to 613)
 Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
 , P.S., Hsiao, C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,
 Seaton, C.L. and Tong, J.C.

TITLE The structure and function of the expressed portion of the wheat
 JOURNAL genomes - 5-15 DAP spike cDNA library
 COMMENT Unpublished (2000)
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105959773
 Fax: 5105959818
 Email: oanderson@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Stratagene SK primer.
 FEATURES
 source Location/Qualifiers
 1..613
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE0903_C03_F05"
 /clone_lib="wheat 5-15 DAP spike cDNA library"
 /tissue_type="Spike"
 /dev_stage="Adult plant"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
 greenhouse. Spikes at 5, 10 and 15 DAP were harvested,
 total RNA and poly(A) RNA were prepared, a cDNA library
 was made, and the cDNA clones were in vivo excised to
 give phagescript phagemids in the TJ Close lab (Choi,
 Close, Fenton) at the University of California,
 Riverside. Plasmid DNA preparations and DNA sequencing
 were performed in the OD Anderson lab (all other authors
)."
 BASE COUNT 178 a 135 c 178 g 122 t
 ORIGIN

Query Match 10.9%; Score 355.2; DB 10; Length 613;
 Best Local Similarity 83.0%; Pred. No. 1.2e-86;
 Matches 405; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 2587 GATACGATGATGACCTGATTGCCAGACTGTTGAAGCTCGGGTGTGAAAAAATAGGAAA 2646
 Db 8 GACATGACGATGACCTGATTCAAACTCATGATGCTTGGGCAATGCAAGAAAGAAC 67

Qy 2647 TAGCTGTTAAGACAGTCAAAACCATCAAGTCTTGGGACTACATGACAGACTCAATAC 2706
 Db 68 TAGCTGTCAAGATACATAAAACCATCAAGTCTTGGGCTGATGAGAGACACAGATTC 127

Qy 2707 CTAAGGGTGTTCGAGCCCATTTGATGATGATTGCCAGCAATGGAACAGAGGATGGCT 2766
 Db 128 CTAAGGATGTTCAAGTGCATTCAGTGTATTGTTGCAACGATGGAACAGAGATGGCT 187

Qy 2767 ATGGCCCAATGAGCAACATGTTGCTGCTCCACGAAAGTCCAGTGAATACCCCG 2826
 Db 188 ATGGCCCAATGAGCAACATGTTGCTGCTCCACGAAAGTCCAGTGAATACCCCG 247

Qy 2827 ACGGCAACCCCGCAAGAACCAAGCAGCTACTGAGTTCGGGAGGTTGAAATG 2886
 Db 248 ACGGCGACCAACCAAGAACCAAGCAGCTACTGAGTTCGGGAGGTTGAAATG 307

Qy 2887 GCAGTCCACCACTGCCATGGCGTTCGCGCTGCGGCTTCCAGCAATAGGGCCCTGC 2946
 Db 308 GCAGTCCACCACTGCCATGGCGTTCGCGCTGCGGCTTCCAGCAATAGGGCCCTGC 367

Qy 2947 TATTGCTAAAGAAATAGCTCCAGCAAGAGGATGATGAGGCTCTGCAACCGGAAATCT 3006
 Db 368 TCTTGTCTCCAGCAACAGCTCCAGCAAGAGGATGATGAGGCTCTGCAACCGGAAATCT 427

Qy 3007 ACGTTCAGCATTTGAGATCTTTGAGTCTGTCGGGATCAAGCTGTTGAGAAAGTGGAGA 3066
 Db 428 ACATCCCTGCGTGGAGATCTTTGAGTCTGTCGGGATCAAGCTGTTGAGAAAGTGGAGA 487


```

QY 3067 CTTGAAAG 3074
Db 488 CCGTAGAG 495

RESULT 15
Bi421952
LOCUS
DEFINITION Bi421952 731 bp mRNA linear EST 16-AUG-2001
EST532618 tomato callus, TAMU Lycopersicon esculentum cDNA clone
CLEC68F1 5' end, mRNA sequence.
ACCESSION Bi421952
VERSION Bi421952.1 GI:15195940
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 731)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Alm,S., Ronning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
Location/Qualifiers
1..731
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC68F1"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/notes="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
BASE COUNT 207 a 155 c 167 g 202 t
ORIGIN

Query Match 10.7% Score 348.8; DB 10; Length 731;
Best Local Similarity 67.6%; Pred. No. 7.5e-85;
Matches 488; Conservative 1; Mismatches 233; Indels 0; Gaps 0;

QY 1793 TCATCGTATTGAAATACACTGCTACCCAGCTTGATGTGCTGATATTGGAAGCCTTTC 1852
Db 5 TAAAGTTATTCCAAATGCAAAAGCAATTCAGCTTGACATTCAGAGTCACGAGTCCTCTTTC 64

QY 1853 AGATCTTGTTCTCAGGTGAGGTGTAATTAGCTTGCTGCTGCTAGTTTTCATGCTGC 1912
Db 65 TAGCTGGATCCGAGAGGTGATGTGTATCATGCTTACTGCTCTCTCATGCTGCCATGGTG 124

QY 1913 CATTCCAGGAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGA 1972
Db 125 TATAGCAAAAGGCATGCATTAGCTGAAGAACAATCTCGTCACGGCTACGTAGCTGTGATGA 184

QY 1973 ATCCATGTCACACITGACCGAAGCTGCCAAAGATGCAGGTGTAACATATCTTTGGAAT 2032
Db 185 TTCTATGTTAAAGCTAGACCAAGATGCAAAATCTGCTGGAATTAATCTTCTGTTGGAAT 244

QY 2033 GGGCCCTAGATCCTGGCATAGATCACTTGATGTCATGAAGATGATTGATGAAGCTATGC 2092
Db 245 GGGCTTGGACCCAGGAATAGATCATATGATGCGGATGAAATGATCAACGAAGCCCATGC 304

```

Search completed: June 1, 2002, 00:56:11
Job time: 7632 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 15:45:23 ; Search time 18.01 Seconds
(without alignments)
1386.061 Million cell updates/sec

Title: US-09-049-304A-122

Perfect score: 5286

Sequence: 1 CARLLGGGKNGPRVRIIV.....VPALEILESSGKIVKVEV 1022

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/1aa/PTCUS_COMB.pep.*

6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	270	5.1	74	1	US-08-474-633A-105
2	263.5	5.0	123	1	US-08-474-633A-104
3	154.5	2.9	370	2	US-08-360-606B-32
4	154	2.9	372	2	US-08-360-606B-31
5	128.5	2.4	382	2	US-08-360-606B-30
6	125.5	2.4	1104	4	US-09-268-347-28
7	125.5	2.4	1104	4	US-09-268-347-34
8	112	2.1	476	4	US-09-171-969-4
9	110.5	2.1	1177	3	US-08-855-160-6
10	110.5	2.1	2254	2	US-08-677-010-3
11	110.5	2.1	2254	2	US-08-790-519-3
12	109.5	2.1	1177	3	US-08-855-160-2
13	109.5	2.1	1177	3	US-08-855-160-8
14	109.5	2.1	1288	1	US-07-727-814B-2
15	109.5	2.1	1288	1	US-08-258-614-2
16	109	2.1	363	1	US-08-096-182A-6
17	109	2.1	363	1	US-08-877-109-6
18	109	2.1	363	3	US-08-798-760-6
19	109	2.1	363	5	PCT-US94-08327-6
20	109	2.1	470	3	US-08-879-565-14
21	109	2.1	623	4	US-09-104-068-4
22	109	2.1	637	4	US-09-104-068-2
23	108.5	2.1	1881	4	US-09-233-086-3
24	107	2.0	1022	3	US-08-772-270A-2
25	107	2.0	1176	1	US-08-446-486-6
26	107	2.0	1176	1	US-08-157-363A-10
27	107	2.0	1176	1	US-08-463-308-6

28 107 2.0 1176 5 PCT-US93-11405A-10 Sequence 10, Appl
29 107 2.0 1176 5 PCT-US93-11405-10 Sequence 10, Appl
30 106 2.0 718 2 US-08-974-565C-7 Sequence 7, Appl
31 106 2.0 718 3 US-09-255-748-7 Sequence 7, Appl
32 106 2.0 1938 4 US-09-514-302-2 Sequence 2, Appl
33 105 2.0 833 4 US-09-514-302-3 Sequence 3, Appl
34 104.5 2.0 1184 6 5254799-6 Patent No. 5254799
35 102.5 1.9 505 1 US-08-068-395A-3 Sequence 3, Appl
36 102.5 1.9 505 1 US-08-464-365-3 Sequence 3, Appl
37 102.5 1.9 893 1 US-07-977-434-4 Sequence 4, Appl
38 102.5 1.9 893 1 US-08-458-819-4 Sequence 4, Appl
39 102.5 1.9 893 4 PCT-US91-07035-4 Sequence 10, Appl
40 102.5 1.9 893 5 PCT-US91-07035-4 Sequence 2, Appl
41 102.5 1.9 1140 4 US-09-220-081-2 Sequence 9, Appl
42 101.5 1.9 1250 3 US-08-938-291A-9 Sequence 10, Appl
43 101 1.9 2366 1 US-08-480-604A-10 Sequence 10, Appl
44 101 1.9 2366 2 US-08-405-496A-10 Sequence 10, Appl
45 101 1.9 2366 4 US-08-915-136-10 Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-474-633A-105
; Sequence 105, Application US/08474633A
; Patent No. 5773691
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
; APPLICANT: COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND
; TITLE OF INVENTION: METHODS FOR INCREASING
; TITLE OF INVENTION: INCREASING THE LYSINE
; TITLE OF INVENTION: AND THREONINE CONTENT
; TITLE OF INVENTION: OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; ADDRESSEE: AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,633A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 105:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-474-633A-105

Query Match 5.1%; Score 270; DB 1; Length 74;
Best Local Similarity 67.6%; Pred. No. 9.9e-19;

```
Matches 50; Conservative 18; Mismatches 6; Indels 0; Gaps 0
QY   948 KHOATLLEFGKVENGRSTTAMALTVGIPAAIGALLLNKNKYOTKGVRPQLPEIYVPALE 1007
      ||||| :||:||| |||||||||:::||:||| |:|||: |||||
Db    1 KHTATLLEFGDKNGQTMTAKTVGIPAAIGALLIEDIKTRGVLRPLEAEVLPALD 60
QY   1008 ILESSGIKLVEKVE 1021
      |||: ||||:|| |
Db    61 ILQAYIGIKLMEKAE 74

RESULT 2
US-08-474-633A-104
; Sequence 104, Application US/08474633A
; Patent No. 5773691
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
; APPLICANT: COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND
; TITLE OF INVENTION: METHODS FOR INCREASING
; TITLE OF INVENTION: INCREASING THE LYSINE
; TITLE OF INVENTION: AND THREONINE CONTENT
; TITLE OF INVENTION: OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; ADDRESSEE: AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474.633A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 123 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-474-633A-104

Query Match          5.0%; Score 263.5; DB 1; Length 123;
Best Local Similarity 51.6%; Pred. No. 1.le-17;
Matches 64; Conservative 17; Mismatches 26; Indels 17; Gaps 4

QY   526 KIGKYNVEYETDITDKGGPK-----ILTGAGRYCRPAAEFPLASYPDICT-----YG 572
      ||||| : : : | : | : ||||| ||||:|||| | : |
Db    1 KIGKVOQ-ENET---KEKPMTKKSGVLILGAGRVRRPAADEFLASVRTISSQQWKTYFG 56
QY   573 VDDHDADQIHIVASLYOKDAEETVDGTENTTATOLDVADIGSLDLVSQVEWISLLPA 632
      | : : ||||| ||||:||||| : | : ||||:| | |||||:|:|||
Db    57 ADSEKTDVHIVASLYLKDAKETVEGISDVRAVLDSDESLSLKYVSQDWDVLSLLPA 116
QY   633 SFHA 636
      | ||
```


QY 18 IIVQPTRRIHQAQYEDACGETSEDLS-----ECGLIIGIKO-PKLOMILSDRAVAFES 71
Db 38 IYVESSQSTFDIKEYEAVGAKIYVPGSWKTAPEKRIIFGLKELPENETFPPLIHEHQFA 97
QY 72 HTHKAOKENMPLDLKILEERVSLFDYELIVGDDKSLAFKFAGRA-----GLIDF---- 123
Db 98 HCYDQAGWODVLRFPFGNGNIIYDLEFLENDQGRVRAAFPGFYAGFAGAAIGVLDWSFKQ 157
QY 124 LHG--LQORYLSLGYSTPFLSLGOSHMYPSLAAKAAVIVVAEEIATFGLPSGICPIVFV 181
Db 158 LNCNTRKTRGEGGELPGVT-----PYPNENELIKDVKIELEKALTKN--GGQYPKCLV 210
QY 182 FTGVGNVSGAQAEIFKLLPHFTFVDAKLELPEIFOARNLSKOSOSTKRVFOLYGCVVTSRDI 241
Db 211 IGALGRCGSGAIDLFRKI-----GIPD-----DNIAKWDM----- 241
QY 242 VSHKDTROPDKGYYAHPEHYTPVHERIAPYASVI-----VNCMYWEKRPPLLNMQD 296
Db 242 -----ETAK-----GPFQEVLDLDFINCIYLSKFIPFPIKKEI 276
QY 297 LOQMETGCPVGVCDITCDIGGSIEFINKSTSIERFFRYDPK--NSYHDM--EAG 351
Db 277 LN---NENRKLTIIVDSAD-----TTNPHDPIVYEIATVNEFTVEVKLDKGF 323
QY 352 GVVCIAVDILPTFESEASOHFNILSRVASLASVQPAELPSYLRACI--AHAGRL 408
Db 324 KLSVCSIDLPLSLPREASEFFA---KDLMPSLLELFRDTSVWVRARQLFKHVARL 379
RESULT 6
US-09-268-347-28
; Sequence 28, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 28
; LENGTH: 1104
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-28

Query Match 2.4%; Score 125.5; DB 4; Length 1104;
Best Local Similarity 22.5%; Pred. No. 0.013;
Matches 111; Conservative 74; Mismatches 193; Indels 115; Gaps 27;
QY 454 INEALDI-IETAGGSPHLVRCVGVQSTDDMSYSELEVAGDDTATL-----DK 499
Db 521 LNLGNLFKSTDGELLNLIKVENDTVTFPKGSGVGVGEGGRATIONGKTTDGLVEASE 580
QY 500 IIDSLSLANEHGGHDAGQGEIHALKIGKVNEYETDVTIDKGGPKILILGAG---RVCR 556
Db 581 LVESLNKLGKVGVDKDGSGELD-----GASNE-----TLVKS GDKV-TLKAGENLKVKQ 629
QY 557 PRAEFILASPDICTYGVDDHDADQIHVIVASLYQKDAEE-----TVDGIENTATQ 607
Db 630 DGTNF-----TYALKDELGT-----VKSVEFKDTANGSNGASTKITKDLTITSANG 676
QY 608 LDVADIGSLDVSOVEVVISLLPASFAHAIAGVCIELKHKMVTASY--VDESMNSLSQA 665
Db 677 AN-----GAATADADKIKVASDGISAG--NKAVNVVSGLAK--FGDANFNPLTSSADNLTKQ 730
QY 666 AKDA--GVITLCEMGLDPGIDHLMMSKMIDEAHARKGIKAFTSYCGGLPSPAAANNPLA 723
Db 731 YDDAYKGLTNLDEKADK-----QTLTVADNTAATVGDLR-----GLGWVISAD---- 774
QY 557 PRAEFILASPDICTYGVDDHDADQIHVIVASLYQKDAEE-----TVDGIENTATQ 607
Db 630 DGTNF-----TYALKDELGT-----VKSVEFKDTANGSNGASTKITKDLTITSANG 676
QY 608 LDVADIGSLDVSOVEVVISLLPASFAHAIAGVCIELKHKMVTASY--VDESMNSLSQA 665
Db 677 AN-----GAATADADKIKVASDGISAG--NKAVNVVSGLAK--FGDANFNPLTSSADNLTKQ 730
QY 666 AKDA--GVITLCEMGLDPGIDHLMMSKMIDEAHARKGIKAFTSYCGGLPSPAAANNPLA 723
Db 731 YDDAYKGLTNLDEKADK-----QTLTVADNTAATVGDLR-----GLGWVISAD---- 774

QY 724 YKESWNPAGALRSKGNPANYKFLGETIHYDGHNLNYESAKRLRLRELPAFALEHLPRNSL 783
Db 775 -----KTTGELNKEYNAQV--RNANEVFKSGNGIHVSGKTVNGRRREITFEL--AKDENAI 826
QY 784 IYGDLYGISK--EASTIYRATRYEGFSEIMVTLSKTGTFDDAANHPLLQDTSRPTVKGFL 841
Db 827 AFG--YG-SKALRDNTVAIGTN-----VYNAKSGAFGDPNY--IEDKAGGSYAFGN 874
QY 842 DELLNNISTINT-----DLIDIEASGGYDDDLIARLLKLGCCCKNKEIAVKTIVTKIFLG 894
Db 875 D---NRITSKNTFVLGNGVNAKYKANGDVEDTETV-----VKDKGKRETTIVTVPKALG 924
QY 895 LHEETOIPKGCSS 907
Db 925 ATVENSIVLGNKS 937
RESULT 7
US-09-268-347-34
; Sequence 34, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 34
; LENGTH: 1104
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-34

Query Match 2.4%; Score 125.5; DB 4; Length 1104;
Best Local Similarity 22.5%; Pred. No. 0.013;
Matches 111; Conservative 74; Mismatches 193; Indels 115; Gaps 27;
QY 454 INEALDI-IETAGGSPHLVRCVGVQSTDDMSYSELEVAGDDTATL-----DK 499
Db 521 LNLGNLFKSTDGELLNLIKVENDTVTFPKGSGVGVGEGGRATIONGKTTDGLVEASE 580
QY 500 IIDSLSLANEHGGHDAGQGEIHALKIGKVNEYETDVTIDKGGPKILILGAG---RVCR 556
Db 581 LVESLNKLGKVGVDKDGSGELD-----GASNE-----TLVKS GDKV-TLKAGENLKVKQ 629
QY 557 PRAEFILASPDICTYGVDDHDADQIHVIVASLYQKDAEE-----TVDGIENTATQ 607
Db 630 DGTNF-----TYALKDELGT-----VKSVEFKDTANGSNGASTKITKDLTITSANG 676
QY 608 LDVADIGSLDVSOVEVVISLLPASFAHAIAGVCIELKHKMVTASY--VDESMNSLSQA 665
Db 677 AN-----GAATADADKIKVASDGISAG--NKAVNVVSGLAK--FGDANFNPLTSSADNLTKQ 730
QY 666 AKDA--GVITLCEMGLDPGIDHLMMSKMIDEAHARKGIKAFTSYCGGLPSPAAANNPLA 723
Db 731 YDDAYKGLTNLDEKADK-----QTLTVADNTAATVGDLR-----GLGWVISAD---- 774
QY 724 YKESWNPAGALRSKGNPANYKFLGETIHYDGHNLNYESAKRLRLRELPAFALEHLPRNSL 783
Db 775 -----KTTGELNKEYNAQV--RNANEVFKSGNGIHVSGKTVNGRRREITFEL--AKDENAI 826
QY 784 IYGDLYGISK--EASTIYRATRYEGFSEIMVTLSKTGTFDDAANHPLLQDTSRPTVKGFL 841
Db 827 AFG--YG-SKALRDNTVAIGTN-----VYNAKSGAFGDPNY--IEDKAGGSYAFGN 874
QY 842 DELLNNISTINT-----DLIDIEASGGYDDDLIARLLKLGCCCKNKEIAVKTIVTKIFLG 894
Db 875 D---NRITSKNTFVLGNGVNAKYKANGDVEDTETV-----VKDKGKRETTIVTVPKALG 924
QY 895 LHEETOIPKGCSS 907


```

FILING DATE: 27-JUL-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/980,128
FILING DATE: 23-NOV-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/808,129
FILING DATE: 13-DEC-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/356,599
FILING DATE: 24-MAY-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 06/904,572
FILING DATE: 05-SEP-1986
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 06/808,129
FILING DATE: 12-DEC-1985
ATTORNEY/AGENT INFORMATION:
NAME: Lloyd, Jeff
REGISTRATION NUMBER: 35,589
REFERENCE/DOCKET NUMBER: M12C1FD#3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (352) 375-8100
TELEFAX: (352) 372-5800
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1177 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-855-160-6

```

Query Match 2.1%; Score 110.5; DB 3; Length 1177;
Best Local Similarity 21.3%; Pred. No. 0.43;
Matches 180. Conservative 100. Mismatches 311. Indels 253.

[illegible]

```

QY 757 -----LYESAKRLR-LRELPFAFALEHLPNRSNLIVGLDYGISKEASTIVRATXRY 805
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 919 QLOADTNIAMIAHAADKRHSIREAYLPSELVPGVNAAIPEELEG-----RIFTAFSLY 972
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 806 EGFSEIMVTLSTGTGFDAAHPLLOOTSRPTYKGLFD-ELLNN-----ISTINTD 854
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 973 DARNVI-----KNGDFN-----NGLSCWNVKGHVDEEQNNQSVLVLPEWAEVSOE 1020
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 855 LDIEAGSGYDDDLIARLLKL-----GCCKNKEIAVTKVTKIKELGLHEETOIPKGCSSP 908
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1021 VRVCPGRGY-----ILRVTAKEYGEGCVTTHEIENNT-DELKFSNCVVEEIIYNNVTVC 1075
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 909 FDVICORMEORMAY-----GHNEQDMVLLHHEVEVEYPDGQPAEKHOATLLEFGKVENGR 963
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1076 NDYTVNQEEVGAYTSRNGYNEAPSV-----PAD--YASVVEEKSYTDGR 1119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 964 STTAMALTVG-----IPAAIGALLLLKNKVQTKGVRPQ--PEIYVPALEILESSGKLV 1017
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1120 RENPCFNRGRDYTPLPGY-----VTKLEYFPETDKVWIEIGETEGTFIV 1167
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1018 EKVE 1021
      : :
Db 1168 DSVE 1171
      : :

RESULT 10
US-08-677-010-3
; Sequence 3, Application US/08677010
; Patent No. 5925805
; GENERAL INFORMATION:
; APPLICANT: Ohlrogge, John B.
; APPLICANT: Roesler, Keith R.
; APPLICANT: Shorrtosh, Basil S.
; TITLE OF INVENTION: Structure and Expression of an
; TITLE OF INVENTION: Arabidopsis Acetyl-coenzyme A Carboxylase Gene
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,010
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 6550-00002CPA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810)641-1600
; TELEFAX: (810)641-0270
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2254 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-677-010-3

```

Query Match 2.1%; Score 110.5; DB 2; Length 2254;
Best Local Similarity 18.6%; Pred. No. 1.4;
Matches 170; Conservative 110; Mismatches 308; Indels 325; Gaps 42;

Db 896 ---GRESHARVIVHSLFE-----EYL-SVELEFN---D 921
QY 575 DHDAOIHVIVASLYKODAEVTD-----GIEN----- 602
Db 922 NMLADVIE-RMQLYKDLKIVDIVLSHQIKNKNKLVRLMEQLVYPNPAAYRDKLIR 980
QY 603 -TTATQLDVADTG-SLSDLVSVQEVVVISLLPASFAAIAAGVCIELKKHMTASYVD----- 656
Db 981 FSTLHNTNYSALAKASQLEOTKL-----SELRSNIARSLSSELEMFTEDEGENMDTPKR 1034
QY 657 -----ESMNLSSQAADAGVTILCEMGLDPGIDHLSMKMIDEAHARKKIKRAFTSY 708
Db 1035 KSAINERIEDIVASLAVEDALV-----GLFDHSDHILQRRVE-----TY 1075
QY 709 CGGLSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRE 768
Db 1076 IRLLYQPVVVKD-SVRMQWHRSGLLASWE----- 1103
QY 769 LPFALEHLPNLSNIYDLYGI-----SKEASTIYRATYRGEFSEIMVTLTKTGFDDAA 824
Db 1104 ---FLEEMERKN-----IGLDDPDTSKGLVEKRKRKRWG---AMVIKSLQFLPSI 1150
QY 825 NHPLLQDTSRPTKGFELDELLAN-----ISTINTDLTIEASGGYDDDLIARLLKLGCCKN 879
Db 1151 ISAAALRETKHNDYETAGAPLSNMHHAIVGINNQMSLLQDSGDEDAQGERVYNKL----- 1205
QY 880 KEIAVTKVTKIFLGHETQIPKGCSSPDV---ICORMEQMAIGHNEQDMVLLHHEV 936
Db 1206 -----AKILK-----EBEVSSSLCSAGVGVISCIQDRGRTPMRHS-----FHWSL 1247
QY 937 EYVEYDGPQAEKH 949
Db 1248 EKQYVVEEPLLRH 1260

RESULT 12

US-08-855-160-2

; Sequence 2, Application US/08855160

; Patent No. 6090931

; GENERAL INFORMATION:

; APPLICANT: Edwards, David L.

; APPLICANT: Herrnstadt, Corinna

; APPLICANT: Wilcox, Edward R.

; APPLICANT: Wong, Siu-yin

; TITLE OF INVENTION: Process for Altering the Host Range of

; TITLE OF INVENTION: Bacillus Thuringiensis Toxins, and No. 6090931el Toxins Produced

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik A

; ADDRESSEE: Professional Association

; STREET: 2421 NW 41st Street, Suite A-1

; CITY: Gainesville

; STATE: Florida

; COUNTRY: USA

; ZIP: 32606-6669

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/855,160

; FILING DATE: 13-MAY-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/580,781

; FILING DATE: 29-DEC-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/420,615

; FILING DATE: 10-APR-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/097,808

; FILING DATE: 27-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/980,128
; FILING DATE: 23-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/808,129
; FILING DATE: 13-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/356,599
; FILING DATE: 24-MAY-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/904,572
; FILING DATE: 05-SEP-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/808,129
; FILING DATE: 12-DEC-1985
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; REGISTRATION NUMBER: 35,589
; REFERENCE/DOCKET NUMBER: M12C1PFD3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (352) 375-8100
; TELEFAX: (352) 372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-855-160-2

Query Match

2.1%; Score 109.5; DB 3; Length 1177;

Best Local Similarity 21.3%; Pred No. 0.54;

Matches 178; Conservative 99; Mismatches 305; Indels 253; Gaps 44;

QY 345 HDDMEGAGVCLAVDIL--PTEFSKEASQHFQGNIL-----SRVVASLASVKQ 389
Db 432 HVSMFRSGFSNSSVSIIRAPTFSQWHRSAEFNNIPSSQITQIPLTKSTNLGSGTSVVKG 491
QY 390 PAEL-PSYLRRACIAHAGRL-----TPLYE-YIPMR-----NTMIDLAPAK---- 429
Db 492 PGFTGGDILRTSPGQISTLRVNTAPLSQRYRVRIRVASTTNLQFHTSIDGRPNQGNF 551
QY 430 -----TNPLPDKKYSTIVLSLGHFLDKFLINEA-LDIE----- 462
Db 552 SATNSSGSLNLOGSPRTVGTTPFNFSNGSSVFTLSAHVFNSSG--NEVYIDRIEFVPAEV 609
QY 463 TAGGSFHLVRCEVGQSTDDMSYSELEVAGDDTAT---LDKIIDSLSLANEHGGHDHDAQ 519
Db 610 TFEAEYDLERAQ--KAVNELEFTSSNQIGLKTVDVTDYHIDQVSNLVECLSFCLDEK--Q 665
QY 520 ETELALKTKG-----VNEY-----ETDVTIDKGGPKIL-----ILG 550
Db 666 ELSEKVHAKRLSDERNLLQDPNPFGRINRQLDRGWRGSTDITI--QGGDDVFKENVYTLIG 724
QY 551 AGRCVCPAAEFASYPDICTYGVDDHDADQIHVIVASLYKQDAEETVDG-----IE 601
Db 725 TFDECYP-----TY-----LYQKIDESKLYATFYQLRGVIE 756
QY 602 NTTATQLDVADIGSLDLSVQEVVISLLPASFAAIAAGVCIELKK-----HMVTASYVDES 658
Db 757 DSQLEIYILIRYNAKHETVN-VPGTGSLLWPLSAQSPI--GKGGEPNRCAPHLEWNPDLDCS 814
QY 659 MSNLSQRAK-----DAQVTILCEMGLDPGIDHLSMKMIDEAHARKKIKRAFTSYC 709
Db 815 CRGCKCAHHSHHPSLDIDVGCTDLNE---DLGVVIFIKITQD--GHARLNLEFEE-- 868
QY 710 GGLSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHN----- 756
Db 869 KPLVGEALARVKRAEK-KWRDKREKLEWETNIVYKEAKESVDALFVNSQYDQLQADTNIA 927

Qy 757 -LYESAKRLR-LRELPAFALEHLPNRSLIYDGLYGISKEASTIVRATXRYEGESEIMVT 814
Db 928 MIHAADKRVHSIREAYLPDSVPGVNAAFEELEG-----RIPTAFSLYDARNVI--- 978
Qy 815 LSKTGFFDAANHLIQLDTSRPTKGFGLD-ELLNN-----ISTINTOLDIEASGGY 863
Db 979 --KNGDFN-----NGLSCNNVKGHVDSBONNORSVLVLPWEAEVSVQVRVCPGRGY 1029
Qy 864 DDDLIARLLK-----GCCNKEIAVTKVTKIRFLGHEETQIPKGCSSPFDVICORME 917
Db 1030 ----ILRVTAKEGYEGECVTIHEIENNT-DELFKFSNVEEIEYPNNTVTCNDYVNOEE 1084
Qy 918 QRMAY-----GHNEQDMVLLHHEVEVEYDPGQPAEKHOATLLEFGKVGNGRSTTAMALTV 972
Db 1085 YGGATSRNGYNAPSV-----PAD--VASVEEKSYYTDGRENPCENR 1128
Qy 973 G-----IPAAIGALLKLNKVKQTKGVRPLQ--PEIYVPALEILESSGKLVKEVE 1021
Db 1129 GYRDYTPLPGV-----VTKELEYFPETDKVMBIEGETEGTFIVDSVE 1171

RESULT 13

US-08-855-160-8
; Sequence 8, Application US/08855160
; Patent No. 6090931

GENERAL INFORMATION:

; APPLICANT: Edwards, David L.
; APPLICANT: Herrnstadt, Corinna
; APPLICANT: Wilcox, Edward R.
; APPLICANT: Wong, Siu-Yin
; TITLE OF INVENTION: Process for Altering the Host Range of
; TITLE OF INVENTION: Bacillus Thuringiensis Toxins, and No. 6090931el Toxins Produced Thereby
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik A
; ADDRESS: Professional Association
; STREET: 2421 NW 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606-6669

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/855,160
; FILING DATE: 13-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/580,781
; FILING DATE: 29-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/420,615
; FILING DATE: 10-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/097,808
; FILING DATE: 27-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/980,128
; FILING DATE: 23-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/808,129
; FILING DATE: 13-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/356,599
; FILING DATE: 24-MAY-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/904,572
; FILING DATE: 05-SEP-1986
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 06/808,129
; FILING DATE: 12-DEC-1985
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; REGISTRATION NUMBER: 35,589
; REFERENCE/DOCKET NUMBER: M12CIFDF3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (352) 375-8100
; TELEFAX: (352) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-855-160-8

Query Match 2.1%; Score 109.5; DB 3; Length 1177;

Best Local Similarity 21.3%; Pred. No. 0.54;
Matches 178; Conservative 99; Mismatches 305; Indels 253; Gaps 44;

Qy 345 HDDMEGAGVVCVLAVDIL--PTEFSKEASQHEFNIL-----SRLVASLASVKQ 389
Db 432 HVSMPFRSGFSNVSIIIRAPTFWQHRSAEFNIIIPSSQITQIPLTKSTNLGSGTSVVKG 491
Qy 390 PAEL-PSYLRRACTIAHAGRL-----TPLYE-YIPMR-----NTMIDLAPAK---- 429
Db 492 PGFTGGDILRRTPSQOISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNF 551
Qy 430 -----TNPLDPKKYSTLVLSGHLGKDFKFLINEA-LDIE----- 462
Db 552 SATMSSGNLQSGSPRTVGFTTFPFNSGSSVFTLSAHVNSG--NEVVIDRIEFVPAEV 609
Qy 463 TAGGSFHLVRCEVGQSTDDMSYSELEVAGDDTAT---LDKIIDSITSLSLANEHGGDHDAQ 519
Db 610 TFEAYDLERAQ--KAVNELFTSSNQIGLTKDVTDYHIDQVSNLVECLSDFCLEK--Q 665
Qy 520 EIELALKTKG-----VNEY-----ETDVTIDKGGPKIL-----ILG 550
Db 666 ELSEKVKHAKRLSDERNLLQDPNFRGINQLDRGRWGSTDITI-QGGDDVFKENYVTLG 724
Qy 551 AGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYOKDAEETVDG-----IE 601
Db 725 TDECYP-----TY-----LYQKIDESKLRATRYQLRGYIE 756
Qy 602 NTTATQLDVADIGSLDLVSQVEVVISLLPASFAHAIAGVCIELKK---HMVTASYVDES 658
Db 757 DSQDLEIVLIRYNKAKHETVN-VPGTGSWPLSAQSPI-GKCGEPNRCAPHLEWNPDLDCS 814
Qy 659 MSNLSQAAK-----DAGVTILCEMLDPCIDHLSMKMIDEAHARKKIKAKFTSVC 709
Db 815 CRDGEKCAHSHHFLSDIDVGCTDLNE---DLGVWVIFKIKTQD-GHARLGNLEFLEE-- 868
Qy 710 GGLPSAAANNPLAYKFSNPNAGALRSRGNPAVYKFLGETIHDVGHN----- 756
Db 869 KPLVGEALARVKAER-KWRDRKLEWETNIVYKEAKESVDALFVNSQYDQLQADNTA 927
Qy 757 -LYESAKRLR-LRELPAFALEHLPNRSLIYDGLYGISKEASTIVRATXRYEGESEIMVT 814
Db 928 MIHAADKRVHSIREAYLPDSVPGVNAAFEELEG-----RIPTAFSLYDARNVI--- 978
Qy 815 LSKTGFFDAANHLIQLDTSRPTKGFGLD-ELLNN-----ISTINTOLDIEASGGY 863
Db 979 --KNGDFN-----NGLSCNNVKGHVDSBONNORSVLVLPWEAEVSVQVRVCPGRGY 1029
Qy 864 DDDLIARLLK-----GCCNKEIAVTKVTKIRFLGHEETQIPKGCSSPFDVICORME 917
Db 1030 ----ILRVTAKEGYEGECVTIHEIENNT-DELFKFSNVEEIEYPNNTVTCNDYVNOEE 1084
Qy 918 QRMAY-----GHNEQDMVLLHHEVEVEYDPGQPAEKHOATLLEFGKVGNGRSTTAMALTV 972
Db 1085 YGGATSRNGYNAPSV-----PAD--VASVEEKSYYTDGRENPCENR 1128
Qy 973 G-----IPAAIGALLKLNKVKQTKGVRPLQ--PEIYVPALEILESSGKLVKEVE 1021
Db 1129 GYRDYTPLPGV-----VTKELEYFPETDKVMBIEGETEGTFIVDSVE 1171

Db 1085 YGAYTSRNGYNEARSV-----PAD--YASVYEKSYTDGRRNPNCEFN 1128

QY 973 G-----IPAAIGALLLNKNVQTKGVRPLO--PEIXVPALEILESSGKILVEKVE 1021

Db 1129 GYRDYTPLPVGY-----VTKLEYFPETDKWVIEIGETEGTFIVDSVE 1171

RESULT 14

US-07-727-814B-2

; Sequence 2, Application US/07727814B

; Patent No. 5213969

; GENERAL INFORMATION:

; APPLICANT: SCHUMACHER, Gunther

; APPLICANT: BURTSCHER, Helmut

; APPLICANT: MOLLERING, Hans

; TITLE OF INVENTION: CLONED N-METHYLHYDANTOINASE

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Armstrong, Nikaido, Marmelstein, Kubovcik &

; STREET: 1725 K Street, N.W., Suite 1000

; CITY: Washington D.C.

; COUNTRY: United States of America

; ZIP: 20006

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/727,814B

; FILING DATE: 19910708

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: DE P 4021571.7

; FILING DATE: 06-JUL-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Murray, Robert B.

; REGISTRATION NUMBER: 22,980

; REFERENCE/DOCKET NUMBER: 911014

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 659-2930

; TELEFAX: (202) 887-0357

; TELE: 440142

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1288 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-07-727-814B-2

Query Match 2.1%; Score 109.5; DB 1; Length 1288;

Best Local Similarity 17.7%; Pred. No. 0.64;

Matches 104; Conservative 93; Mismatches 236; Indels 153; Gaps 22;

QY 347 DMEAGVVCVLAVIDLPTFSKESQHFNGILSRVLASLASVQPAELPSYLRACIAHAG 406

Db 152 ELKTAGVQIAVCLLHSLYNPEHQRIEIVNE-----EFP-----EAYLSLS 195

QY 407 RLTPLYEYIPRNMIDIAKTNPLDPKKYSTVLVLSGHL-FDKFLINEALDIETAG 465

Db 196 ETVPLYREYERFSTALN---AYGVPRVSRYLHRLQEQAEHLGYQREIL-----LMQSSG 247

QY 466 GSFHLVRCVCGVSTDDMS-----YSELEVGADETATLDKIIDSITSLSANEHGDH 515

Db 248 GMPVGEARAKRPVTLMMSGPVGGLCGMMAKQSGFENVVTLD-----ICGTS 296

QY 516 DAGQEIETALKI-----GKNVEYETDV-----TIDKGGPKILLILGAGRVCRPAAEFLAS 564

Db 297 DIGVAYQGLRMRHLLDTKIGHQAWPMVDITDIGAGGSIAYVDAGGVFRVGPQSAGA 356

QY 565 YPDICYG-----VDDHDAQI-----HVIIVASLYQKD---AEETVDGIENITATQLD 609

Db 357 VPGVCYGRGCTPTSTDAQVLLGRMRPDRILAGSLDMDLDRARAAMQGLADKLGMSIE 416

QY 610 VADIGSLSDLVSOVEVVISLIPASFAHAATAGVCIEIKHMTASYVDESMSNLSQAQDA 669

Db 417 EAALGAL-----QIQKF-----GMTQAIEQNSVRRGYDPDRFTLV--AAGGA 456

QY 670 GVTILCEMLDPCGDHLSMKMIDEAHARKGKTKAFTSYCGGLPSPAANNPLAYKFSWN 729

Db 457 GALFACEIAAELEVPHVL-----VPAHPGIAGIIGLLATDQYEFVATN--RFSFAFR 507

QY 730 PACALRSKGNPAVYKFLGETIHVDGHNLYESAKRRLRLRELPAPAFALHPLNRNSLIYCDLY 789

Db 508 DAAVIQAS-----YEQLERERNAQLDA---EEVP-----533

QY 790 GISKEASTIYRATRYEG-----FSEIMVTLSTKGTFFDAAHPLLOQTSRPTTKGFL 841

Db 534 --AERKIVMLRDARYEGQYEIFRVVPEGPVTTAFLDQAEAFHDAHFEEYGHREFKGGT 591

QY 842 DELLNNISTINTDLDIEASGGVDDDLIARLLKLGCCCKNEIAVKTV 887

Db 592 VEVIN-----IRVEARVMDLPTPEATOSGSLENALVETRPV 629

RESULT 15

US-08-258-614-2

; Sequence 2, Application US/08258614

; Patent No. 5432070

; GENERAL INFORMATION:

; APPLICANT: SCHUMACHER, Gunther

; APPLICANT: BURTSCHER, Helmut

; APPLICANT: MOLLERING, Hans

; TITLE OF INVENTION: CLONED N-METHYLHYDANTOINASE

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram

; STREET: 655 Fifteenth Street N.W. Suite 330

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005-5701

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/258,614

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/021,856

; FILING DATE: 24-FEB-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: DE P 40 21 571.7

; FILING DATE: 06-JUL-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/727,814

; FILING DATE: 08-JUL-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Kitts, Monica C.

; REGISTRATION NUMBER: 36,105

; REFERENCE/DOCKET NUMBER: P564-3007

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)638-5000

; TELEFAX: (202)638-4810

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1288 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-258-614-2

```
Query Match      2.18; Score 109.5; DB 1; Length 1288;
Best Local Similarity 17.74; Pred. No. 0.64;
Matches 104; Conservative 93; Mismatches 236; Indels 153; Gaps 22;

QY 347 DMGAGVVCCLAVDILPTFEFSKASQHFNTILSRIVASLASVKQPAELPSYLRACIAHAG 406
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 ELKTAGVQAIAVCLLHSLYLNPEHQRIEIVNE-----EFP-----EAYLSLS 195

QY 407 RLTPLYEIPMRMTMDLAPAKTNPLPKKYSTLVSLSGHL-FDKFLINEALDIITETAG 465
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 196 EIVPLRYEYERFESTALN---AYVGRVSRYLHRLQEAENLGVQREIL-----LMQSSG 247

QY 466 GSFHLVRCEVGQSTDDMS-----YSELEVGADDTATLDKIIDSLSLANEHGGDH 515
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 248 GWPVPIGEAAKRPVTLMSGPGVGLIGGMWAAKQSGFENVVTLT-----IGGTS 296

QY 516 DAGOEIELALKI-----GKNEYETDV-----TIDKGGPKILILGAGRVCRPAAEFLAS 564
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 DIGVAYQOGLMRHLLDTKIGDHOAMVPMVDIDTIGAGGSIAYVDAGGVFRVGPQOSAGA 356

QY 565 YPDICTYV-----VDDHDADOI-----HVIVASLYQKD---AETVPGIENTTATOLD 609
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 VPGPVYGRGCTEPTSTDAQVLLGRMPDRILAGSLDMDLDRARAAMQGLADKLGMSIE 416

QY 610 VADIGSLDLVSQVEVVISLLPASFHAATAGVCIELKHHMTASYVDESMSNLQAOKDA 569
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 417 EAALGAL-----QIQKF-----GMTQATEQNSVRRGYDPRDFTLV--AAGGA 456

QY 670 GVTILCEMLDPCIDHLMNKKMIDEAHARKGKIKAFSTYCGGLPSPAAANNPLAYKFSWN 729
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 GALTACEIAAELEVPVHL-----VPAHPGIIAGIGLLATDEQYEFVATN--RESFAFR 507

QY 730 PAGALRSKGNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLY 789
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 508 DAAVIQAS-----YEQLERERNAQLDA---EVP-----EVP----- 533

QY 790 GTSKEASTIYRATXRYEG-----FSEIMVTLTKTGFDAANHPLQDTSRPTYKGFL 841
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 534 --AERRKIWLDRARYEGGYEIRFVVPVPTTAWLDQAEAAAFHDAHFEEYGHFRKGGT 591

QY 842 DELLNNISTINTDLDIEASGGYDDDLIARLLKLGCKKNKEIAVKTV 887
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 592 VEVIN-----IRVEARVMDELPTPTATQSGSLENALVETRPV 629
```

Search completed: May 31, 2002, 15:48:06
Job time: 163 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 15:44:38 ; Search time 39.13 Seconds
(without alignments)
2901.033 Million cell updates/sec

Title: US-09-049-304a-122

Perfect score: 5286

Sequence: 1 CARLLGGKNGKPRVNRIV.....VPALEILSSGKIKVEKVT 1022

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

	1:	2:	3:	4:	5:	6:	7:	8:	9:	10:	11:	12:	13:	14:	15:	16:	17:	18:	19:	20:	21:	22:
	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5284	100.0	1022	20	Maize lysine ketog
2	3416	64.6	1064	20	Arabidopsis lysine
3	3411	64.5	1064	21	Arabidopsis thalia
4	3316	62.7	640	20	Maize lysine ketog
5	3190.5	60.4	969	21	Arabidopsis thalia
6	3089.5	58.4	948	21	Arabidopsis thalia
7	2498.5	47.3	784	20	Arabidopsis thalia
8	1480	28.0	482	20	Soybean lysine ket
9	1466.5	27.7	928	22	Arabidopsis saccha
10	1466.5	27.7	972	22	Drosophila melanog
11	996	18.8	239	20	Rice lysine ketog

12	360	6.8	83	20	AAW87765	wheat lysine ketog
13	270	5.1	74	19	AAW60533	Protein homologous
14	270	5.1	74	20	AAW87758	Arabidopsis lysine
15	263.5	5.0	123	19	AAW60532	Protein homologous
16	263.5	5.0	123	20	AAW87757	Arabidopsis lysine
17	181.5	3.4	145	22	AAW84624	Human protein sequ
18	179	3.4	42	20	AAW87764	Rice lysine ketog
19	144.5	2.7	382	17	AAW00495	Candida albicans s
20	130.5	2.5	605	11	AAW03923	Bacillus megateriu
21	125.5	2.4	1104	21	AAW83856	Haemophilus influe
22	125.5	2.4	1104	21	AAW83856	Haemophilus influe
23	120	2.3	491	15	AAW60342	Partial human lipi
24	120	2.3	493	20	AAW49556	Human cholesterol
25	120	2.3	493	20	AAW49556	Human cholesterol
26	120	2.3	493	20	AAW49556	Human cholesterol
27	119.5	2.3	1349	22	AAW79343	Human polypeptide
28	119.5	2.3	1349	22	AAW79343	Human polypeptide
29	119.5	2.3	1349	22	AAW79343	Human polypeptide
30	119	2.3	774	20	AAW37454	Protein involved 1
31	117.5	2.2	1194	22	AAW83599	Human polypeptide
32	117.5	2.2	1194	22	AAW83599	Human polypeptide
33	117	2.2	1118	22	AAW8264	Rice magnesium che
34	116	2.2	632	22	AAW35319	Enterococcus faeca
35	116	2.2	923	22	ABB59510	Drosophila melanog
36	114.5	2.2	3177	22	AAW40303	Human polypeptide
37	112.5	2.1	1243	20	AAW35286	C. pneumoniae prot
38	112	2.1	476	17	AAW06127	Human cholesterol
39	112	2.1	476	18	AAW46446	Human mature chole
40	112	2.1	476	20	AAW02466	Human cholesterol
41	111.5	2.1	531	22	ABW60201	Drosophila melanog
42	111.5	2.1	1023	16	AAW76991	LhaA (low homology
43	111.5	2.1	1421	21	AAG39649	Arabidopsis thalia
44	111.5	2.1	1499	21	AAG39648	Arabidopsis thalia
45	111.5	2.1	1655	21	AAG39647	Arabidopsis thalia

ALIGNMENTS

RESULT 1

ID	AAW87761	standard; Protein; 1022 AA.
XX	AAW87761;	
AC		
DT	29-MAR-1999 (first entry)	
XX		
DE	Maize lysine ketoglutarate reductase.	
XX		
DE	Lysine ketoglutarate reductase; saccharopine dehydrogenase;	
KW	transgenic plant; seed; maize.	
XX		
OS	Zea mays.	
XX		
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 803	/note= "encoded by YTT"
XX		
XX	WO9842831-A2.	
PN		
PD	01-OCT-1998.	
XX		
XX	27-MAR-1998;	98WO-US06051.
XX		
PR	27-MAR-1997;	97US-0824627.
XX		
PA	(DUFO) DU PONT DE NEMOURS & CO E I.	
XX		
PI	Epelbaum SU, Falco SC, McDevitt RE;	
XX		
DR	WPI; 1999-045139/04.	
DR	N-PSDB; AAV99563.	
XX		

PT Nucleic acids and chimeric genes for increasing seed lysine content
PT - comprise sequence encoding all or part of lysine ketoglutarate
PT reductase, useful to improve nutritional quality of seeds from
PT transformed plants
XX
PS Claim 2; Page 189-191; 231pp; English.
XX This is the amino acid sequence of a maize near full-length
CC lysine ketoglutarate reductase (LKR) or saccharopine dehydrogenase
CC (SDH). It was deduced from cDNA (see AAV99563) obtained from
CC developing seed mRNA by PCR and RACE. Isolated nucleic acids
CC comprising sequences encoding all or part of plant LKR enzymes are
CC new. Also claimed are: (i) a chimeric gene comprising the
CC fragment (or a subfragment) operably linked to a seed specific
CC regulator, where the chimeric gene reduces LKR activity in plant
CC seeds transformed with it; (2) plant cells and seeds in which LKR
CC activity is reduced due to a mutation in the gene encoding LKR or
CC transformation with the chimeric gene; (3) a nucleic acid fragment
CC comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC synthase (DHPS) substantially insensitive to lysine inhibition is
CC operably linked to a plant chloroplast transit sequence and to a
CC plant seed-specific regulatory sequence; (4) plants comprising in
CC genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC from (3). The chimeric genes can be used to produce plant cells
CC and seeds with reduced LKR activity, especially in Arabidopsis,
CC corn, soybean, rapeseed, wheat and rice (claimed). LKR is
CC important in controlling free lysine accumulation in plant seeds.
CC LKR activity reduction may be achieved by cloning the claimed
CC fragment, preparing a chimeric gene for cosuppression of LKR,
CC expression of antisense RNA for LKR, and transforming plants with
CC the chimeric gene.
XX
SQ Sequence 1022 AA;

Query Match 100.0%; Score 5284; DB 20; Length 1022;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1022; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CARLLGGGKNGPRVRIIVQPTSTRHHDAQYEDAGCEISELSECGLIIGIKQPKLOM 60
DB 1 CARLLGGGKNGPRVRIIVQPTSTRHHDAQYEDAGCEISELSECGLIIGIKQPKLOM 60
QY 61 ILSDRAVAFSTHKAQENPLDKILERVSLFDYELIVGDDGKRSKSLAFGRFAGRGL 120
DB 61 ILSDRAVAFSTHKAQENPLDKILERVSLFDYELIVGDDGKRSKSLAFGRFAGRGL 120
QY 121 IDFLHGLQRYLSLGYSTPFLSLGQSHMYPSLAAKAAVIVAAEETATFGLPSGICPIVF 180
DB 121 IDFLHGLQRYLSLGYSTPFLSLGQSHMYPSLAAKAAVIVAAEETATFGLPSGICPIVF 180
QY 181 VFTGVGNVSGQAQETFKLLPHTFVDAEKLPEIFQARNLSKOSQSTKRVFOLYGCVTSRD 240
DB 181 VFTGVGNVSGQAQETFKLLPHTFVDAEKLPEIFQARNLSKOSQSTKRVFOLYGCVTSRD 240
QY 241 IVSHKDPTRQFDKGDYAHPEHYTPVFERHATPYASVIVNCWYWEKRPPLLNMDQLQQL 300
DB 241 IVSHKDPTRQFDKGDYAHPEHYTPVFERHATPYASVIVNCWYWEKRPPLLNMDQLQQL 300
QY 301 METGCPVLGVCDITCDIGSGSTEFINKSTIERPFRYPDPKSNYSYHDDMEGAGVVCVAVDI 360
DB 301 METGCPVLGVCDITCDIGSGSTEFINKSTIERPFRYPDPKSNYSYHDDMEGAGVVCVAVDI 360
QY 361 LPTEFSKASQHFQNLRLSVASLASVKQPAELFSLYRRACIAHAGRLTPLYEYIPRMN 420
DB 361 LPTEFSKASQHFQNLRLSVASLASVKQPAELFSLYRRACIAHAGRLTPLYEYIPRMN 420
QY 421 TMIDLAPAKTNPLPKKYSTVLSLGHLPDFKPLINEALDIETAGGSPHLVRCFVGSTGD 480
DB 421 TMIDLAPAKTNPLPKKYSTVLSLGHLPDFKPLINEALDIETAGGSPHLVRCFVGSTGD 480
QY 481 DMSYSELEVAGDATTATLDKIIDSLSLANEHGGDHDAQEIEALKIKGVANEYETDVTID 540

DB 481 DMSYSELEVAGDATTATLDKIIDSLSLANEHGGDHDAQEIEALKIKGVANEYETDVTID 540
QY 541 KGGPKLILILGAGRCRPAAEFLASYPDICVGVDDHDAQIHHVIVASLYOKDAEETVDGI 600
DB 541 KGGPKLILILGAGRCRPAAEFLASYPDICVGVDDHDAQIHHVIVASLYOKDAEETVDGI 600
QY 601 ENTATOLDVADIGSLSDLYSQVEVWISLSPASFHAAIAGVCIELKKHMVTASYVDESMS 660
DB 601 ENTATOLDVADIGSLSDLYSQVEVWISLSPASFHAAIAGVCIELKKHMVTASYVDESMS 660
QY 661 NLSAAKADAGVTILCEMGLDPGIDHLSMKMIDEAHARKGKIKAFSTYCGGLSPAAANN 720
DB 661 NLSAAKADAGVTILCEMGLDPGIDHLSMKMIDEAHARKGKIKAFSTYCGGLSPAAANN 720
QY 721 PLAYKFSWNPAGALRSKGNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNR 780
DB 721 PLAYKFSWNPAGALRSKGNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNR 780
QY 781 NSLIYGDLYGTSKEASTIYRATXRYEGFSEIMVTLSTGTFDAAHPLLOOTSRPTYKGF 840
DB 781 NSLIYGDLYGTSKEASTIYRATXRYEGFSEIMVTLSTGTFDAAHPLLOOTSRPTYKGF 840
QY 841 LDELLNNISTINTDLDTASGYYDDDLIARLLKLGCCNKKEIAYKTVTKIKFLGLHEETQ 900
DB 841 LDELLNNISTINTDLDTASGYYDDDLIARLLKLGCCNKKEIAYKTVTKIKFLGLHEETQ 900
QY 901 IPKGCSSPFDVICQRMEOARMAYGNEODMVLHHEVEVEYDPGQPAEKHQTILFEKVE 960
DB 901 IPKGCSSPFDVICQRMEOARMAYGNEODMVLHHEVEVEYDPGQPAEKHQTILFEKVE 960
QY 961 NGRSTTAMALVGPAAATGALLILKNKVQTKGVRPLQPELYVPALEILESSGKILVEKV 1020
DB 961 NGRSTTAMALVGPAAATGALLILKNKVQTKGVRPLQPELYVPALEILESSGKILVEKV 1020
QY 1021 ET 1022
DB 1021 ET 1022
RESULT 2
AAW87759
ID AAW87759 standard; Protein; 1064 AA.
XX
AC AAW87759;
XX
DT 29-MAR-1999 (first entry)
XX
DE Arabidopsis lysine ketoglutarate reductase.
XX
KW Lysine ketoglutarate reductase; saccharopine dehydrogenase;
XX Transgenic plant; seed.
XX
OS Arabidopsis thaliana cv. Landsberg erecta.
XX
PN W09842831-A2.
XX
PD 01-OCT-1998.
XX
PF 27-MAR-1998; 98WO-US06051.
XX
PR 27-MAR-1997; 97US-0824627.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Epelbaum SU, Falco SC, McDevitt RE;
XX
DR WPI; 1999-045139/04.
XX
DR N-PSDB; AAV99559.
XX
PT Nucleic acids and chimeric genes for increasing seed lysine content
PT - comprise sequence encoding all or part of lysine ketoglutarate
PT reductase, useful to improve nutritional quality of seeds from
PT reductase, useful to improve nutritional quality of seeds from

transformed plants

Claim 2; Page 175-178; 231pp; English.

This is the amino acid sequence of Arabidopsis thaliana lysine ketoglutarate reductase (LKR)/saccharopine dehydrogenase (SDH), as deduced from a gene (see AAV9599) isolated from a CD4-8 Landsberg erecta genomic library. The protein lacks an N-terminal targeting sequence suggesting that the lysine degradative pathway in Arabidopsis is located in the plant cell cytosol. Isolated nucleic acid fragments comprising sequences encoding all or part of plant LKR polypeptides are new. Also claimed are: (1) a chimeric gene comprising the fragment (or a subfragment) operably linked to a suitable seed specific regulator, where the chimeric gene reduces LKR activity in plant seeds transformed with it; (2) plant cells and seeds in which LKR activity is reduced due to a mutation in the gene encoding LKR or transformation with the chimeric gene; (3) a nucleic acid fragment comprising: (i) chimeric gene above, and (ii) a second chimeric gene, in which a nucleic acid fragment encoding dihydrodipicolinic acid synthase (DHDDS) substantially insensitive to lysine inhibition is operably linked to a plant chloroplast transit sequence and to a plant seed-specific regulatory sequence; (4) plants comprising in genome (i) and (ii), especially as fragment of (2); and (5) seeds and seeds with reduced LKR activity, especially in Arabidopsis, corn, soybean, rapeseed, wheat and rice (claimed). LKR is important in controlling free lysine accumulation in plant seeds. LKR activity reduction may be achieved by cloning the claimed fragment, preparing a chimeric gene for co-suppression of LKR, expression of antisense RNA for LKR, and transforming plants with the chimeric gene.

Sequence 1064 AA;

Query Match 64.6%; Score 3416; DB 20; Length 1064;
Best Local Similarity 63.6%; Pred. No. 5.2e-295;
Matches 668; Conservative 152; Mismatches 178; Indels 52; Gaps 13;

QY 1 CARLLGGKGNPRVNRIVPSTRRHHDAQYEDAGCEISEDLSECLGIIGIKOPKLOM 60
DB 38 car-llhggkgrtgsrivrsvsakrllhhdalyehvgceisdldsdglllgkqpelem 96
QY 61 ILSDRAYAFFSHTHKAENPLLDKILEERVSLFDFYELIVGDGKRSAPFAGKAGRL 120
DB 97 ilperayaffshthkaenplldkileervslfdfydelivgdgkrlafgkyagr 156
QY 121 IDFLHGLQRYLSGYSTPFLSLGQSHMPSLAATAAIVVAAEVIATFGLPSGICPIVF 180
DB 157 vdfihglqrylslygstpfslsgasymyslaaakaavsvgeiasqglp.gicplvf 216
QY 181 VFTGVNVSOGAQTETKLLPHTFVDAKLEPIF-OARNLSKQSQSTRKRVQLYGCVVTSR 239
DB 217 vftgvnvsogaeifkllphtfvepsklpelvfkkqisngistkrvvygvciltsq 276
QY 240 DIVSHKDPTRQDFKGDYAHPEHTVPFHERIAPYASVIVNCMYEKRFPPLNMDLOLQ 299
DB 277 dmvehkdpksfkdadyahpehynpvhkispysvlnvncmyekrfpcilstkgld 336
QY 300 LMETGCPVLGVCDTCDIGGSIEFINKSTIERPFYDPFSKNSYHDDMEGAGWCLAVD 359
DB 337 ltkkgplvgldctcdiggsiefvnratlidsppfrfnpsnnsyvdmdgdgvlcmavd 396
QY 360 ILPTFESKEASQFNTLSRLVASLAVSKQAPLPSYLRACIAHAGRTPLVEYIPMR 419
DB 397 ilptefekasqfngdlisgfvslasmeisdiphlkracisyrgetlsyeyipmr 456
QY 420 NTMIDLAPAKTNP-----LPDKKYSTVLSLGHLPFKFLNEALDIETAGGS 467
DB 457 -----ksnpeeaqndliangvssqtrfnllvslshlfdkflnealmeaaggs 507
QY 468 FHLVRCEVGQSTDDMSYSELEVCAADTATLDKIIDLSTSLAN---EHGDDHDAQGEIELA 524

DB 508 fhllakcelggsadaesylevsgaddkrvldqldsltrlanpnedyisphreanki--s 565
QY 525 LKIGKVNVEYTDVTDKGGPK-----LILGAGRVCPAAEFASYDPDICT-----Y 571
DB 566 lkigkvqg-enei---kekpentkksvglilgagrvcpaadflasvrtissqqwykyf 621
QY 572 GYDDHDADQIHIVIVASLYQKDAEETVDGTEENTATQLOVDADTGSLSLVSQVEVVISLLP 631
DB 622 gadseektvhrivaslylkdkatevegslveavrlvdsdesllkyvsqdvvlslp 681
QY 632 ASFHAAIAGVCIELKKHMTVASYVDESNLSQAADAGVTILCEMGLDPGIDHLMMSKM 691
DB 682 aschavvakctielkhlvtasyvddetshhekaksagilgemglpoidhmmamkm 741
QY 692 IDEAHARKGKTAKFTSYCGGLPSPAAANPLAYKFSWNPAGALSGKNPAAVYKFLGETIH 751
DB 742 indahikkkvksfscycgglpspaaanpplaykfswnpagairagqnpakykangdlih 801
QY 752 VDHNLVESAARLRLRELPAFALHLPNRNLSLIYGDLYGSKAEASTYVATXRYVEGFSEI 811
DB 802 vdgknlvdsaafrvplnpafalecfpnrdslvygehyeseattifrgtlrlyegfsmi 861
QY 812 MVTLSKGTGFDAANHPILQDTSRTTYKGFDELNNISTINTDLDIEASGYYDDDLIARL 871
DB 862 matlskgtgfdaanhpilqdstsrttykfgdelnnistintdldieasgyydddlia 916
QY 872 LKLGCCRNKEIAVTVTKIRFLGLHEETQTPKGCSPDFVICQPMQEMOAMAYGHNEQDWL 931
DB 917 iklg--hsketaakaaktivflgineerevpslcksvfdatcylmeeklaysgneqdmv 974
QY 932 LHHEVEVEYDGPQAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLKNKVOTK 991
DB 975 lhhevevefleskriekhtatllefgdikngqttamaktvgipaaigalvltedkitr 1034
QY 992 GVRPLPEIYVPALETLLESSGKLVKVE 1021
DB 1035 gvrplpeaevlpaldilqaygiklmeake 1064

RESULT 3
AAG31238
ID AAG31238 standard; Protein; 1064 AA.
XX AAG31238;
XX AAG31238;
DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 37481.
XX Arabidopsis thaliana
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.

us-09-049-304a-122.rag

Tue Jun 4 15:23:12 2002

PR	21-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145085.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132863.	PR	22-JUL-1999;	99US-0145192.
PR	11-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145145.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134219.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134221.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145913.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145918.
PR	19-MAY-1999;	99US-0134941.	PR	27-JUL-1999;	99US-0145919.
PR	20-MAY-1999;	99US-0135124.	PR	28-JUL-1999;	99US-0145951.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146388.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0146389.
PR	27-MAY-1999;	99US-0136392.	PR	03-AUG-1999;	99US-0147038.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147204.
PR	01-JUN-1999;	99US-0137222.	PR	04-AUG-1999;	99US-0147302.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147192.
PR	04-JUN-1999;	99US-0137502.	PR	05-AUG-1999;	99US-0147260.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147303.
PR	08-JUN-1999;	99US-0138094.	PR	06-AUG-1999;	99US-0147416.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138847.	PR	09-AUG-1999;	99US-0147935.
PR	14-JUN-1999;	99US-0139119.	PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139452.	PR	11-AUG-1999;	99US-0148319.
PR	17-JUN-1999;	99US-0139453.	PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148585.
PR	18-JUN-1999;	99US-0139454.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139455.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139456.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139457.	PR	18-AUG-1999;	99US-0149436.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139463.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139750.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139751.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151338.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	02-JUL-1999;	99US-0142055.	PR	20-SEP-1999;	99US-0154779.
PR	06-JUL-1999;	99US-0142390.	PR	22-SEP-1999;	99US-0155139.
PR	08-JUL-1999;	99US-0142803.	PR	23-SEP-1999;	99US-0155486.
PR	09-JUL-1999;	99US-0142920.	PR	24-SEP-1999;	99US-0155659.
PR	12-JUL-1999;	99US-0142977.	PR	28-SEP-1999;	99US-0156458.
PR	13-JUL-1999;	99US-0143542.	PR	29-SEP-1999;	99US-0156596.
PR	14-JUL-1999;	99US-0143624.	PR	04-OCT-1999;	99US-0157117.
PR	15-JUL-1999;	99US-0144005.	PR	05-OCT-1999;	99US-0157753.
PR	16-JUL-1999;	99US-0144085.	PR	06-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	07-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	08-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	12-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.

PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0139584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	26-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 64.5%; Score 3411; DB 21; Length 1064;			
Best Local Similarity 63.6%; Pred. No. 1.5e-294;			
Matches 668; Conservative 151; Mismatches 179; Indels 52; Gaps 13;			
Qy	1	CARLLGGKNGPRVNRITVOPSTRHHDAQYEDAGCEISEDLSECCLIIGIKOPKLOM 60	
Db	38	car-llhggkrtglsrvlpsakrihhdaldyedvgeisddldsglllgikqpelem 96	
Qy	61	ILSDRAYAFFSHTHKAQENMPLDKLLEERVSLFDVELIVGDDGKSLAPGKFAAGL 120	
Db	97	ilperayaffshthkaqemmpldkllservtlcdyvelivgdhgrllragkyagrll 156	
Qy	121	IDFLHGLQORVLSLGYSTPFLSLQCSHMYPSLAACAIVVAEEIATFGLPSGICPIVF 180	
Db	157	vdfllhlgqrklilgystpflslqcsmysslaaakaavsvgeiasqgplgicplvf 216	
Qy	181	VFTGVGNVSOQAQIFKLLPHTFYDAKLPEIF-OARNLSKQSOSTRVQLYGCVTSR 239	
Db	217	vftgtgnvslgaqefklphtfvepsklpelvfkdqglsqngistkrvygvyciitsq 276	
Qy	240	DIVSHKQPTQFDKDYAHPEHTVPHERAPIASYIVNCMYWEKRFPPPLNMDQLQ 299	
Db	277	dmvehkdpkskfdkadyahpehynpvfhekispysvlvncmywekrfcpcllstkqlq 336	
Qy	300	LMETGCPVLVGDITDTCIGGSIERFINKSTIERPFFRYDPKSNYHDDMGAGVVC LAVD 359	
Db	337	ltkkgplvglditcdidggsiefvnratlldspffrnpnsnyyddmdgdgvlcmavd 396	
Qy	360	ILPTEFSKEASQHFNGNLSRLVASVKQAPALPSYLRACIAHAGRLTPLYEYIPMR 419	
Db	397	ilptefakeasqhfngdlilsgfvslasmtelsdlpahlkrcacisyrgetlsyeyipmr 456	
Qy	420	NTMIDLAPAKTNP-----LPDKXYSTVLSLGHLPDKFLINEALDIETAGGS 467	
Db	457	-----ksnpeeaqdnllangvsqrtnilvslgshlfkflineaalmdieaaggs 507	
Qy	468	FHLVRCEVGQSTODMSYSELEVAGDATTATDKTIIDSLTSLAN--EHGGDHDAGQIEIA 524	
Db	508	fhlakcelgsadaesylevagdadrldqldidltrlanpnedyisphreanki--s 565	
Qy	525	LKTKGVNEYETDVIKGGPK-----ILTAGRCVRPAAEFLASYPDICT-----Y 571	
Db	566	lkigkvqq-enei---kexpemtkksqgvlilgagrcvrpaadflasvrtissqwyktyf 621	
Qy	572	GVNDHDADQIHTVVASLYOKDAEETVDGIENTATATOLDVADTGSLSDLVSQVEVVISLPL 631	
Db	622	gadseektdhvliavsllykdketvegisdveavrlvdsvdesllykysqvdvvlslip 681	
Qy	632	ASFHAAIAGVCIELKHHMYTASYVDESMSNLSQAARKDAGVTILCEMGLDPDGIHLSMKM 691	

Db	682	aschavvaktclelkkhlvtasyvddetsmihkakagesitilgemglpogldhmmankm 741	
Qy	692	IDEAHARKGKTKAFYSYCGGLPSPAANPLAYKFSWMPAGALRSGKNPAYSFKLGETIH 751	
Db	742	indahikkvkvkstysycgglpspaanplaykfswnpagairagnpakyskngdlih 801	
Qy	752	VDGHNLYESAKRLRLRELPFALEHLNPNRSLIYGDLYGSKSEASTIYRATXRREGEFSEI 811	
Db	802	vdgknlvdsaarfrvnpalpafalecfnrdslvgehygeiseattifrgtirvegfsml 861	
Qy	812	MYTLSKGTFFDAANPLLDQTSRPTYKGFLELLNNISTINTDLDIASGGYDDDLIARL 871	
Db	862	matlskigfdeanqvlstgkrlitfgallsnlnk-----dadneseplag-eeieskri 916	
Qy	872	LKLGCKNKEIAVKTIKFGLGHEETOIPKGCSSPFDVICQRMFORMAYGHNEQDWL 931	
Db	917	iklg--hsketaakaaktivflgfneerevpslcksvdatcyimeeklaysgneqdmvl 974	
Qy	932	LHHEVEVEYPDGQPAEKHQATLLEFGKVENGRSTAMALTVGIPAAIGALLLNKNVOTK 991	
Db	975	lhhevevfleskriehtatllefgdikngqttamaktvgipaaigalliedkiktr 1034	
Qy	992	GVIRPQPEIYVPALEILESIGKLVKEVE 1021	
Db	1035	gvirpleaeavylpaldilgagvikmekae 1064	
RESULT 4			
AAW87762			
ID	AAW87762	standard; Protein; 640 AA.	
XX	AAW87762;		
AC	AAW87762;		
DT	29-MAR-1999	(first entry)	
XX	Maize lysine ketoglutarate reductase partial sequence.		
DE	Lysine ketoglutarate reductase; saccharopine dehydrogenase;		
XX	transgenic plant; seed; maize.		
KW	Zea mays.		
OS	WO9842831-A2.		
PN	01-OCT-1998.		
XX	27-MAR-1998; 98WO-US06051.		
PF	27-MAR-1997; 97US-0824627.		
XX	(DUPO) DU PONT DE NEMOURS & CO E I.		
PA	Epelbaum SU, Falco SC, McDevitt RE;		
XX	WPI; 1999-045139/04.		
PI	N-PSDB; AAV99564.		
XX	Nucleic acids and chimeric genes for increasing seed lysine content		
DR	- comprise sequence encoding all or part of lysine ketoglutarate		
XX	reductase, useful to improve nutritional quality of seeds from		
PT	transformed plants		
XX	Claim 2; Page 193-195; 231pp; English.		
PS	This is the amino acid sequence of a maize lysine ketoglutarate		
XX	reductase (LKR) or saccharopine dehydrogenase (SDH) partial		
CC	polyptide. It was deduced from cDNA (see AAV99564) obtained from		
CC	developing seed mRNA by PCR and RACE. Isolated nucleic acids		
CC	comprising sequences encoding all or part of plant LKR enzymes are		
CC	new. Also claimed are: (i) a chimeric gene comprising the		
CC	fragment (or a subfragment) operably linked to a seed specific		
CC	regulator, where the chimeric gene reduces LKR activity in plant		

CC seeds transformed with it; (2) plant cells and seeds in which LKR
CC activity is reduced due to a mutation in the gene encoding LKR or
CC transformation with the chimeric gene; (3) a nucleic acid fragment
CC comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC synthase (DHPS) substantially insensitive to lysine inhibition is
CC operably linked to a plant chloroplast transit sequence and to a
CC plant seed-specific regulatory sequence; (4) plants comprising in
CC genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC from (3). The chimeric genes can be used to produce plant cells
CC and seeds with reduced LKR activity, especially in Arabidopsis,
CC corn, soybean, rapeseed, wheat and rice (claimed). LKR is
CC important in controlling free lysine accumulation in plant seeds.
CC LKR activity reduction may be achieved by cloning the claimed
CC fragment, preparing a chimeric gene for cosuppression of LKR,
CC expression of antisense RNA for LKR, and transforming plants with
CC the chimeric gene.
XX
SQ Sequence 640 AA;

Query Match 62.7%; Score 3316; DB 20; Length 640;
Best Local Similarity 100.0%; Pred. No. 1.9e-286;
Matches 640; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CARLLGGCKNGPRNRIIVQSTRIIHHDAOYEDAGCEISEDLSECGLIIGIKQPKLQM 60
DB 1 CARLLGGCKNGPRNRIIVQSTRIIHHDAOYEDAGCEISEDLSECGLIIGIKQPKLQM 60
QY 61 ILSDRAYAFSSHAKAKENMPLDKILEERVSLDYELIVGDDGKRSIAFGKFAGRAGL 120
DB 61 ILSDRAYAFSSHAKAKENMPLDKILEERVSLDYELIVGDDGKRSIAFGKFAGRAGL 120
QY 121 IDPLHGLGORYSLGYSTPFLSGQSHMYPSLAAKAAVIVVAEEIATFGLPSGICPIVF 180
DB 121 IDPLHGLGORYSLGYSTPFLSGQSHMYPSLAAKAAVIVVAEEIATFGLPSGICPIVF 180
QY 181 VFTGVGNVSGAQEIPKLLPHTFVDAEKLPETFPQARNLSKQSQSTKRVFQLYGCVVTSRD 240
DB 181 VFTGVGNVSGAQEIPKLLPHTFVDAEKLPETFPQARNLSKQSQSTKRVFQLYGCVVTSRD 240
QY 241 IVSHKDPTRQFDKGYAHPEHTVPFHERIAPYASVIVNCMYWEKRRFPPLNMDLOQL 300
DB 241 IVSHKDPTRQFDKGYAHPEHTVPFHERIAPYASVIVNCMYWEKRRFPPLNMDLOQL 300
QY 301 METGCPVLGVCDITCDIGGIEFINKSTIERPFFRYDPSKNSYHDDMEGAGVVCVLAVDI 360
DB 301 METGCPVLGVCDITCDIGGIEFINKSTIERPFFRYDPSKNSYHDDMEGAGVVCVLAVDI 360
QY 361 LPTEFSKEASQHFGNILSRILVASLAVKQPAELPSYLRACIAHAGRLTPLYEYIPMRN 420
DB 361 LPTEFSKEASQHFGNILSRILVASLAVKQPAELPSYLRACIAHAGRLTPLYEYIPMRN 420
QY 421 TMIDLAPAKTNPLPKKYSLVSLGHLDFKELINEALDIIETAGGSFHLVRCVEGQST 480
DB 421 TMIDLAPAKTNPLPKKYSLVSLGHLDFKELINEALDIIETAGGSFHLVRCVEGQST 480
QY 481 DMSYSELEVAGDPTATLDKIDSITSLSANEGHGDHAGQAGIEALKTGKVNNEYETDVTID 540
DB 481 DMSYSELEVAGDPTATLDKIDSITSLSANEGHGDHAGQAGIEALKTGKVNNEYETDVTID 540
QY 541 KGGPKIILGAGRCVRAAEFLASYPDICTYGVDDHDADQIHVIVASLYOKDAEETVDGI 600
DB 541 KGGPKIILGAGRCVRAAEFLASYPDICTYGVDDHDADQIHVIVASLYOKDAEETVDGI 600
QY 601 ENTATOLDVADIGSLSDLSQVFWISLLPASPHAAIAG 640
DB 601 ENTATOLDVADIGSLSDLSQVFWISLLPASPHAAIAG 640

RESULT 5
AAG31239
ID AAG31239 standard; Protein: 969 AA.

XX
AC AAG31239;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37482.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127452.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 11-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.

```
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144336.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 05-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147319.
PR 06-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154038.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 60.4%; Score 3190.5; DB 21; Length 969;
Best Local Similarity 63.3%; Pred. No. 5.8e-275;
Matches 627; Conservative 141; Mismatches 172; Indels 51; Gaps 12;

Qy 60 MILSDRAYAFETHKAKQENMPLLDKILBERVSLFDYELIVGDDGKRSLAFGRFAGRAG 119
Db 1 milperayafishthkkaqenmplldkilservtlcdyeligdhgkrllafgkyagrag 60

Qy 120 LIDFLHGLGQRYLSLGYSTPFLSLGQSHMYPYSLAAKAAVIVAAETATATFCLPSGICPIV 179
Db 61 lvdflhglgrkllgystptflslgasymyslaaakaavlsveelasgplgicplv 120

Qy 180 FVTGTGNVSGQAEIFKLLPHTFVDAEKLPETIF-QARNLSKQSOSTKRVFQLYGCVVTS 238
Db 121 fvtgtgnvslgaqeifkllphtfvepsklpelfvkdkglsqngistckrvyvygciits 180

Qy 239 RDIVSHKDPTRQFDKGDYIAHPEHYTPVFHERTAPYASVINCMYWEKRPPLINMDQLQ 298
Db 181 qdmvenkdpksfdkadyahpelnynpvhfekispyslvnvcmywekrfpcllstkqlq 240
```

QY	299	QLMETGCPVLGVCDITCDIGGSIERPFYDPSKNSYHDDMEGAGVVCCLAV	358	XX	25-FEB-2000;	2000EP-0301439.
Db	241	dltkkglplvgicdtcdigsgietvrnatlidsfrfnnsnyddmgdgvicmav	300	XX	25-FEB-1999;	99US-0121825.
QY	359	DILPTEFSEKASOHRGNILSRVLASVAKOAPAEPSYLRRACIAHAGRLPLYEIPRM	418	PR	05-MAR-1999;	99US-0123180.
Db	301	dlilpcefakaesqhfgdillsrgvslamsteisdlpahkkracisrgelstelyeiprm	360	PR	09-MAR-1999;	99US-0123548.
QY	419	RNTMIDLAPAKTNP-----LDPKYSTVLVSGHLFDFKFLNEALDIETAGG	466	PR	23-MAR-1999;	99US-0125788.
Db	361	r-----ksnpeeaqdnliangvssqrtnilvslghlfdkflnealdmeaagg	411	PR	25-MAR-1999;	99US-0126264.
QY	467	SFLHVRCEVQSTDDMSYSELEVAGDDATLTKIIDSLSLAN---EKGDDHAGQEL	523	PR	29-MAR-1999;	99US-0126785.
Db	412	sflhakcelggadaesyselevagddkrvidqldsltrlanpnedylsphreanki--	469	PR	01-APR-1999;	99US-0127462.
QY	524	ALKIKGVNEYEDVITDKGGPK-----ILILGAGRCVRPAEFLASYPDICT	570	PR	06-APR-1999;	99US-0128234.
Db	470	sikigkvq-enei---kekpmtkksgvlligagrcrpaadflasvrtissqqwykty	525	PR	08-APR-1999;	99US-0128714.
QY	571	YGVDDHDADQIHVVIVASLYOKDAETVDGIENTTATQLDVADIGSLDSLVQVEVVISLL	630	PR	16-APR-1999;	99US-0129845.
Db	526	fgadseektvhwivasylykdakvetegsdveavridvsdesllkyvsqdvvisll	585	PR	19-APR-1999;	99US-0130077.
QY	631	PASFHAAIAGVCTELKHKHVTASYVDESMNSLSQAADAGVVTILCEMGLDPGIDHLSMK	690	PR	21-APR-1999;	99US-0130449.
Db	586	pachavaktctelckhlvtasyvddetsmlhekaksagtilgemglpdiidhmmamk	645	PR	23-APR-1999;	99US-0130510.
QY	691	MIDEAHARKGKIKAFTSYCGGLPSPAANNPLAYFSPWNPAGALSGKNPAVYKFLGETI	750	PR	28-APR-1999;	99US-0130891.
Db	646	mindahikkgkvksfscyggipspaaannplaykfswnpagaigragqpakyksngdii	705	PR	28-APR-1999;	99US-0131449.
QY	751	HVDGHNLYSARLRLRELPALALEHLPNRNSLIYGLDYGISKEASTIVRATRYEGFGE	810	PR	30-APR-1999;	99US-0132048.
Db	706	hvdgknlvysaarfrvnpilpalecfpnrslvgyehygieaseattifglrlryegfsm	765	PR	04-MAY-1999;	99US-0132486.
QY	811	IMVTLSTGTFDFAANPLPLQTSRPTKGLFDLLELNINSTITDLDIEASGGVDDDLIAR	870	PR	06-MAY-1999;	99US-0132487.
Db	766	imatlsklgffgdeanqvistgkrtifgallsnlnk----dadneseplag-eeeslkr	820	PR	07-MAY-1999;	99US-0132863.
QY	871	LLKLGCKNKETAVKTVKTKFLGLHEETQIPKGCSSPFDVICORMEQRWAYCHNEQDMV	930	PR	11-MAY-1999;	99US-0134250.
Db	821	ilklg--hsketaakaaktviflgfneerevpslcksvfidatcyimeeklaysgneqdmv	878	PR	14-MAY-1999;	99US-0134218.
QY	931	LLHHEVEVEYDGPAPKQKQATLLEKGVENGSRSTTAMALTGVIPAAIGALLILKNKQVT	990	PR	14-MAY-1999;	99US-0134221.
Db	879	llhhevefleskriekhtatillefgdikngqgttamaktvgipaagalliedkikt	938	PR	18-MAY-1999;	99US-0134370.
QY	991	KGVRPLQPEIYYPALLESIGIKLVKVE 1021		PR	18-MAY-1999;	99US-0134768.
Db	939	rgvlrpleaevyypaldilqaygikimekae 969		PR	20-MAY-1999;	99US-0134941.
RESULT	6			PR	20-MAY-1999;	99US-0135124.
AAG31240				PR	21-MAY-1999;	99US-0135353.
ID	AAG31240 standard; Protein; 948 AA.			PR	24-MAY-1999;	99US-0135629.
XX				PR	25-MAY-1999;	99US-0136021.
AC				PR	27-MAY-1999;	99US-0136392.
XX				PR	28-MAY-1999;	99US-0136782.
DT	17-OCT-2000 (first entry)			PR	01-JUN-1999;	99US-0137222.
XX				PR	03-JUN-1999;	99US-0137528.
DE				PR	04-JUN-1999;	99US-0137502.
XX				PR	07-JUN-1999;	99US-0137724.
KW				PR	08-JUN-1999;	99US-0138094.
KW				PR	10-JUN-1999;	99US-0138540.
KW				PR	10-JUN-1999;	99US-0138847.
OS				PR	14-JUN-1999;	99US-0139119.
XX				PR	16-JUN-1999;	99US-0139452.
XX				PR	16-JUN-1999;	99US-0139453.
XX				PR	17-JUN-1999;	99US-0139492.
XX				PR	18-JUN-1999;	99US-0139454.
XX				PR	18-JUN-1999;	99US-0139455.
XX				PR	18-JUN-1999;	99US-0139456.
XX				PR	18-JUN-1999;	99US-0139457.
XX				PR	18-JUN-1999;	99US-0139458.
XX				PR	18-JUN-1999;	99US-0139459.
XX				PR	18-JUN-1999;	99US-0139460.
XX				PR	18-JUN-1999;	99US-0139461.
XX				PR	18-JUN-1999;	99US-0139462.
XX				PR	18-JUN-1999;	99US-0139463.
XX				PR	18-JUN-1999;	99US-0139750.
XX				PR	18-JUN-1999;	99US-0139753.
XX				PR	21-JUN-1999;	99US-0139817.
XX				PR	22-JUN-1999;	99US-0139899.
XX				PR	23-JUN-1999;	99US-0140353.
XX				PR	23-JUN-1999;	99US-0140354.
XX				PR	24-JUN-1999;	99US-0140695.
XX				PR	28-JUN-1999;	99US-0140823.
XX				PR	29-JUN-1999;	99US-0140991.
XX				PR	30-JUN-1999;	99US-0141287.
XX				PR	01-JUL-1999;	99US-0141842.
XX				PR	01-JUL-1999;	99US-0142154.
XX				PR	02-JUL-1999;	99US-0142055.
XX				PR	06-JUL-1999;	99US-0142390.

Arabidopsis thaliana protein fragment SEQ ID NO: 37483.
Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
Arabidopsis thaliana.
EP1033405-A2.
06-SEP-2000.

PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145313.
PR	27-JUL-1999;	99US-0145318.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145918.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	04-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148365.
PR	16-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	16-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	23-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 58.4%; Score 3089.5; DB 21; Length 948;		
Best Local Similarity 62.7%; Pred. No. 5.7e-266;		
Matches 608; Conservative 140; Mismatches 171; Indels 51; Gaps 12;		
QY	81	MPLDKILEERVSLFDYELIVGDDGKRSLAFGKRFAGRAGLIDFLHGLGORYLSLGYSTPF 140
Db	1	mplldkilservtlcdyelyvgdhgkrlafgkyagraglvdfhlg9gklllgyucpf 60
QY	141	LSLQSHMYPSLAAAKAAIVVAEEIATFGLPSGICPIVFVFTGCVNVSOGAQBIFKLLP 200
Db	61	lslgasymyslaaakaavlsvglsgqslplglcplvftgtgnvslgaqefklp 120
QY	201	HTFVDAEKLPEIF-QARNLSKQSOSTKRVFQLYGCVVTSRDIVSHKOPTQFQDKGDIYAH 259
Db	121	htfvepsklpelvfkkgisqngistkrvyqvygvcitssqdmvehkpsksfdkadyah 180
QY	260	PEHYTPVFERIAPYASVIVNCMYWEKRFPPLLNMDQLQOLMETGCPVLVGVCDTCTDIGG 319
Db	181	pehynpvfhekispysvlnvcmywekrfpclstklqldtkkglplvlgldctcdig 240
QY	320	SIEFINKSTIERPFFRYDFPSKNSYHDDMEGAGVVCVLAVDILPTEFSKEASQHFENLSR 379
Db	241	siefvnratlidsppfrfnpsnnsyddmgdgvvlcmavdilpctefakeasqhfegdlisg 300
QY	380	LVASLASVKQPAELPSYLRACIAHAGRLPLVEYIPRMTMTIDLPAKTNP----- 432
Db	301	fvgslasmtelsdplahlkracisyrgetlsyleyipmr-----ksnpeeaqdni 351
QY	433	-----LPDKKYSTLVSLSGHLFDKFLINEALDIETAGSGFHLVRCVWGOSTDDMSYSEL 487
Db	352	iangvssqrtnilvslsghlfdkflinealdmleaaggshfakelgsgadaesysel 411
QY	488	EVGADDTATLDKIIDSLSLAN---EHGGDHDAGQETELAKIGKVNEYETDVTIDKGGP 544
Db	412	evgaddkrvidqiidsltrlanpnedyisphreankl--slkigkvqg-enel---kekp 465

Db 543 leclpnrnsllgydlgit-eastifrgtlyryegfseimgtltslrlsfnnahsllmqg 601
Qy 834 RPTVKGFLDELNNISINTDLDIEASGYDDLIARLLKLGCCCKNEIAVTKTKFL 893
Db 602 rptfkfelfllkvgnpndellg-----endimegillghckdgrtmetaktifl 656
Qy 894 GLHEETOIPKGCSSPFDFVICORMEORMAYGHNEDVLLHHEVEYEPDGPAAEKHOATL 953
Db 657 gllqgteipascksaafvacfmeerlsystekdmvllhheveieydpdqitekhratl 716
Qy 954 LEFGKVENGRSTTAMALTVGIPAAIGALLLNKRVQTKGVIRPLOPEIYVPALEILESSG 1013
Db 717 lefgktidekttamaltvgipaavagalllntkigtgrvrlrpiepeynpaldieayg 776
Qy 1014 IKLVKE 1021
Db 777 ikliekte 784

RESULT 8
AAW87766
ID AAW87766 standard; Protein; 482 AA.
XX
AC AAW87766;
XX
DT 29-MAR-1999 (first entry)
XX
DE Arabidopsis saccharopine dehydrogenase.
XX
KW Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW transgenic plant; seed.
XX
OS Arabidopsis thaliana.
XX
PN W09842831-A2.
XX
PD 01-OCT-1998.
XX
PF 27-MAR-1998; 98WO-US060051.
XX
PR 27-MAR-1997; 97US-0824627.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Epelbaum SU, Falco SC, McDevitt RE;
XX
DR WPI; 1999-045139/04.
XX
DR N-PSDB; AAV99568.
XX

Nucleic acids and chimeric genes for increasing seed lysine content
- comprise sequence encoding all or part of lysine ketoglutarate
reductase, useful to improve nutritional quality of seeds from
transformed plants

Claim 2; Page 201-202; 23lpp; English.

This is the amino acid sequence of the saccharopine dehydrogenase
(SDH) domain of the lysine ketoglutarate reductase (LKR)/SDH protein
of Arabidopsis thaliana. Nucleic acid fragments comprising
sequences encoding all or part of plant LKR polypeptides are new.
Also claimed are: (1) a chimeric gene comprising the fragment (or a
subfragment) operably linked to a suitable seed specific regulator,
where the chimeric gene reduces LKR activity in plant seeds
transformed with it; (2) plant cells and seeds in which LKR
activity is reduced due to a mutation in the gene encoding LKR or
transformation with the chimeric gene; (3) a nucleic acid fragment
comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
in which a nucleic acid fragment encoding dihydrodipicolinic acid
synthase (DHDS) substantially insensitive to lysine inhibition is
operably linked to a plant chloroplast transit sequence and to a
plant seed-specific regulatory sequence; (4) plants comprising in
genome (i) and (ii), especially as fragment of (2); and (5) seeds
from (3). The chimeric genes can be used to produce plant cells

CC and seeds with reduced LKR activity, especially in Arabidopsis,
CC corn, soybean, rapeseed, wheat and rice (claimed). LKR is
CC important in controlling free lysine accumulation in plant seeds.
CC LKR activity reduction may be achieved by cloning the claimed
CC fragment, preparing a chimeric gene for cosuppression of LKR,
CC expression of antisense RNA for LKR, and transforming plants with
CC the chimeric gene.
XX
SQ Sequence 482 AA;

Query Match 28.0%; Score 1480; DB 20; Length 482;
Best Local Similarity 60.0%; Pred. No. 6, 7e-123;
Matches 290; Conservative 78; Mismatches 101; Indels 14; Gaps 4;

Qy 546 ILILGAGRCRPAAEFLASYPDICT-----YGVDDHDADQIHVIVASLYOKDAETVD 598
Db 7 vlllgagrvcrrpaadflasvrtissqwyktyfygdsseektdvhlvaslylkdakctve 66
Qy 599 GIENITATQLDVADIGSLSDLSQVEVVISLLPASFAHAIAGVCIELKHHMTASYVDES 658
Db 67 glsdveavrlvdsdesllkyvsqvvvlslpaschavavaktciellkhlvtasyvdde 126
Qy 659 MSNLSQAAKADAGVTTLCWGLDPGIDHLSMKMIDEAHARKGKIKAFKTSYCGGLSPAAA 718
Db 127 tsmihkakagagittlgengldpgldhmmamkmindahikgkvksfscyegglspaaa 186
Qy 719 NNPLAYKFSWNPAGALRSGKNPVPYKFLGETIHVDGHNLYESAKRLRLRELPAPALEHLP 778
Db 187 nnplaykfswnpagairagqpakykngdilhvdgknllydsaaarfrvnpipafalecfp 246
Qy 779 NRNSLIYGDLYGISKEASTIYRATRYEGFSEIMVTLSKTGFFDAANPLLDQTSRPTYK 838
Db 247 nrdslyvgehygieseattifrgtlyryegfsmatlsklgffdsanqvistgkrtifg 306
Qy 839 GFLDELLNNISINTDLDIEASGGYDDLIARLLKLGCCCKNEIAVTKTKFLGLHDEE 898
Db 307 allsnlnk----dadneseplag-eeiskriiklg--hsketaakaaktviflgnee 359
Qy 899 TQIPKGCSSPFDFVICORMEORMAYGHNEDVLLHHEVEYEPDGPAAEKHOATLLEFGK 958
Db 360 revpslcksvdctylmeeklaysgneqdmvllhhevefleskriekhtatillefgd 419
Qy 959 VENGESTTAMALTVGIPAAIGALLLNKRVQTKGVIRPLOPEIYVPALEILESSGIKLV 1018
Db 420 ikngtqtamaktvgipaalgvlviediktrgvrlrpieaeavypaldilqaygikime 479
Qy 1019 KVE 1021
Db 480 kae 482

RESULT 9
AAE05075
ID AAE05075 standard; Protein; 928 AA.
XX
AC AAE05075;
XX
DT 12-SEP-2001 (first entry)
XX
DE Drosophila melanogaster LKR/SDH protein.
XX
KW Fruitfly; lysine-ketoglutarate reductase; LKR; pesticide; therapy;
KW saccharopine dehydrogenase; SDH; invertebrate enzyme; drug screening;
KW peptidyl-alpha-hydroxyglycine alpha-amidating lyase; PAL;
XX genetic modification; lysine catabolism disorder.
OS Drosophila melanogaster.
XX
FH Key Location/Qualifiers
FT Domain 23..474
FT /note= "Alanine dehydrogenase/pyridine nucleotide
transhydrogenase domain"

FT	Domain	485..498	
FT	/note= "Lactate/malate dehydrogenase domain"		
XX	WO200149856-A2.		
XX	12-JUL-2001.		
XX			
PD			
XX			
XX	05-JAN-2001; 2001WO-US00360.		
XX			
XX	06-JAN-2000; 2000US-0174973.		
XX	29-FEB-2000; 2000US-0185844.		
PR	22-MAR-2000; 2000US-0191189.		
XX			
PA	(GENO-) GENOPTERA LLC.		
XX			
XX	Stout TU, Keegan KP, Ebens AJ;		
PI	WPI; 2001-441885/47.		
XX	N-PSDB; AAD08977.		
DR			
DR			
XX			
PPT	Novel invertebrate enzymes and nucleic acid encoding the enzyme useful		
PPT	as targets for pesticides and to identify compounds that have utility		
PPT	as therapeutics or pesticides -		
XX			
XX	Claim 15; Page 49-51; 54pp; English.		
XX			
CC	The invention relates to invertebrate enzymes and nucleic acids,		
CC	including lysine-ketoglutarate reductase/saccharopline dehydrogenase		
CC	(LKR/SDH) and peptidyl-alpha-hydroxyglycine alpha-amidating lyase (PAL)		
CC	from Drosophila melanogaster, designated dMLKR/SDH, dMPAL and dMPAL2.		
CC	The enzymes are useful for detecting a candidate compound especially a		
CC	putative pesticidal or pharmaceutical agent that interacts with an		
CC	invertebrate enzyme or its fragment. The identified compound is useful		
CC	for controlling a pest, by reducing viability of the pest. Insect or		
CC	worm genetically modified to express the enzyme are useful for studying		
CC	invertebrate enzyme activity and for identification and screening of		
CC	pesticide targets directed to components of a pathway involving a		
CC	subject protein. Nucleic acids encoding the invertebrate enzymes or their		
CC	fragments are useful as biopesticides. The enzymes and their DNA are		
CC	useful for genetically modifying metazoan invertebrate organisms, such		
CC	as insects and worms or cultured cells, resulting in expression or		
CC	mis-expression of the protein. The organisms or cultured cells are useful		
CC	for screening candidate compounds for the ability to modulate expression		
CC	or activity of a subject nucleic acid or protein, and thus are useful in		
CC	identification of new drug targets, therapeutic agents, diagnostics and		
CC	prognostics useful in treatment of disorders associated with abnormal		
CC	enzyme levels and/or function, e.g. disorders associated with lysine		
CC	catabolism. The present sequence is Drosophila melanogaster LKR/SDH		
CC	(dMLKR/SDH) protein. LKR/SDH is a principal enzyme in the catabolism of		
CC	lysine.		
XX			
XX	Sequence 928 AA;		
XX			
Query Match 27.7%; Score 1466.5; DB 22; Length 928;			
Best Local Similarity 33.4%; Pred. No. 3.2e-121;			
Matches 343; Conservative 183; Mismatches 318; Indels 183; Gaps 22;			
QY	17	RIIVQPSSTRIRHDAQVEDAGCEISDELSCGLITIGIKQPKLQMLSDRAVAFESHTHKA 76	
Db	54	kviqpnsrraypmqymqagahiqedisasvifgkvqpidaligpktycfishtika 113	
QY	77	QKENMPLDLKILERSVLSFLYELIVGDDGRKSLAFGFRAGRLIDFLHGLGQRYLSUGY 136	
Db	114	qesnmplldaillekkirliyderrgarqvafgkyagvagnvnlhglgtrllalgh 173	
QY	137	STPFLSLGQSHMYPSPSLAAAAAVTVAEEIATFCLPSGICPIVFVFYGVGNVSGOAOEIF 196	
Db	174	htpfmhigpahnynasmarqairdcyelslgnmpksigplttvfgsgnvsgagevf 233	
QY	197	KLLPHTFVDAEKUPEITFQARNLSKQSOTRRVFOLYGCVVTSRDVIIVSHKDPTRDFDKGY 256	
Db	234	selpievvpemlrkvaehgnqk-----lygcevsrdhlerre-ggqfdakey 282	

XX Epeibaum SU, Falco SC, McDevitt RE;
 PI WPI: 1999-045139/04.
 DR N-PSDB; AAV99565.
 XX
 XX Nucleic acids and chimeric genes for increasing seed lysine content
 PT - comprise sequence encoding all or part of lysine ketoglutarate
 PT reductase, useful to improve nutritional quality of seeds from
 PT transformed plants
 XX
 XX Claim 2: Page 196-197: 231pp; English.
 PS
 XX This is the amino acid sequence of a rice lysine ketoglutarate
 CC reductase (LKR) or saccharopine dehydrogenase (SDH) partial
 CC polypeptide. It was deduced from cDNA (see AAV99565) obtained from
 CC an EST database search. Isolated nucleic acid fragments
 CC comprising sequences encoding all or part of plant LKR enzymes
 CC are new. Also claimed are: (1) a chimeric gene comprising the
 CC fragment (or a subfragment) operably linked to a seed specific
 CC regulator, where the chimeric gene reduces LKR activity in plant
 CC seeds transformed with it; (2) plant cells and seeds in which LKR
 CC activity is reduced due to a mutation in the gene encoding LKR or
 CC transformation with the chimeric gene; (3) a nucleic acid fragment
 CC comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
 CC in which a nucleic acid fragment encoding dihydrodipicolinic acid
 CC synthase (DHDS) substantially insensitive to lysine inhibition is
 CC operably linked to a plant chloroplast transit sequence and to a
 CC plant seed-specific regulatory sequence; (4) plants comprising in
 CC genome (i) and (ii), especially as fragment of (2); and (5) seeds
 CC from (3). The chimeric genes can be used to produce plant cells
 CC and seeds with reduced LKR activity, especially in Arabidopsis,
 CC corn, soybean, rapeseed, wheat and rice (claimed). LKR is
 CC important in controlling free lysine accumulation in plant seeds.
 CC LKR activity reduction may be achieved by cloning the claimed
 CC fragment, preparing a chimeric gene for cosuppression of LKR,
 CC expression of antisense RNA for LKR, and transforming plants with
 CC the chimeric gene.
 XX
 SQ Sequence 239 AA:
 Query Match 18.8%; Score 996; DB 20; Length 239;
 Best Local Similarity 83.2%; Pred. No. 3.1e-80;
 Matches 188; Conservative 21; Mismatches 17; Indels 0; Gaps 0;
 QY 616 LSDLVQVEVWISLLPASFAAAGVCIELKHMVTASYVDESMNSLSQAADGVTILC 675
 Db 4 lsnlvqvevrvllpasfhaaiarvciemkhlvtasydesmsklegagvtiic 63
 QY 676 ENGLDPEIDHLSMKKVIDEHAARKGKIKAPTSYCGGLPSPAANNPLAYKFSWNPAGALR 735
 Db 64 emglpdxhmsmkmideahsrkgkikstscgglpspasannplaykfwspagair 123
 QY 736 SGKNPAYVFLGFTIHVDGHNLYESAKRLRLPFALEHLPRNSLNYGLDLYGSKEA 795
 Db 124 agnpayvfhgeliivdgdklyesakrlrlxelpafalehlprnslmygdlyglskea 183
 QY 796 STIYRATXRYEGFSEIWMVTLTKGFFDAANHPLLODTSRPTKGF 841
 Db 184 stvratlyegfneimattfakigffdaashp1lqtrtptyxdl 229

RESULT 12

AAW87765

ID AAW87765 standard; Protein; 83 AA.

XX AAW87765;

AC AAW87765;

XX 29-MAR-1999 (first entry)

DT Wheat lysine ketoglutarate reductase partial sequence.

XX

XX

XX

KW Lysine ketoglutarate reductase; saccharopine dehydrogenase;
 KW transgenic plant; seed; wheat
 XX Triticum aestivum.
 OS
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 58 /note= "encoded by NTG"
 FT Misc-difference 62 /note= "encoded by ATN"
 FT
 XX
 PN W09842831-A2.
 XX
 XX 01-OCT-1998.
 XX
 PF 27-MAR-1998; 98WO-US06051.
 XX
 PR 27-MAR-1997; 97US-0824627.
 XX
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 XX Epeibaum SU, Falco SC, McDevitt RE;
 XX
 XX WPI: 1999-045139/04.
 DR N-PSDB; AAV99567.
 DR
 XX Nucleic acids and chimeric genes for increasing seed lysine content
 PT - comprise sequence encoding all or part of lysine ketoglutarate
 PT reductase, useful to improve nutritional quality of seeds from
 PT transformed plants
 XX
 XX Claim 2: Page 199; 231pp; English.
 PS
 XX This is the amino acid sequence of a wheat lysine ketoglutarate
 CC reductase (LKR) or saccharopine dehydrogenase (SDH) partial
 CC polypeptide. It was deduced from cDNA (see AAV99567) obtained from
 CC an EST database search. Isolated nucleic acid fragments
 CC comprising sequences encoding all or part of plant LKR enzymes
 CC are new. Also claimed are: (1) a chimeric gene comprising the
 CC fragment (or a subfragment) operably linked to a seed specific
 CC regulator, where the chimeric gene reduces LKR activity in plant
 CC seeds transformed with it; (2) plant cells and seeds in which LKR
 CC activity is reduced due to a mutation in the gene encoding LKR or
 CC transformation with the chimeric gene; (3) a nucleic acid fragment
 CC comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
 CC in which a nucleic acid fragment encoding dihydrodipicolinic acid
 CC synthase (DHDS) substantially insensitive to lysine inhibition is
 CC operably linked to a plant chloroplast transit sequence and to a
 CC plant seed-specific regulatory sequence; (4) plants comprising in
 CC genome (i) and (ii), especially as fragment of (2); and (5) seeds
 CC from (3). The chimeric genes can be used to produce plant cells
 CC and seeds with reduced LKR activity, especially in Arabidopsis,
 CC corn, soybean, rapeseed, wheat and rice (claimed). LKR is
 CC important in controlling free lysine accumulation in plant seeds.
 CC LKR activity reduction may be achieved by cloning the claimed
 CC fragment, preparing a chimeric gene for cosuppression of LKR,
 CC expression of antisense RNA for LKR, and transforming plants with
 CC the chimeric gene.
 XX
 SQ Sequence 83 AA;

Query Match 6.8%; Score 360; DB 20; Length 83;

Best Local Similarity 83.1%; Pred. No. 2.9e-24;

Matches 69; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 940 YPDGQPAEKHQATLLLEFGKVENGRSTAMALTVGIPAAIGALLILKKNKVTGKVRPLQP 999

Db 1 ypdgdpkqhqtlllefgkntenrpttamaltvgvpaagallllqknkvrkvrpexp 60

QY 1000 EIVVPALEITLESSGKILVEKVT 1022

Db 61 exypaleileasgiklieret 83

RESULT 13
AAW60533
ID AAW60533 standard; Protein: 74 AA.
XX
XX AAW60533;
AC
XX 25-AUG-1998 (first entry)
DT
XX Protein homologous to fungal saccharopine dehydrogenases.
DE
XX Dihydrodipicolinic acid synthase; DHDPs; chimeric gene;
KW lysine inhibition; plant chloroplast transit sequence;
KW plant seed-specific regulatory sequence; transgenic plant;
KW increased lysine level; corn; Zea mays; soybean; glycine max.
XX
XX Arabidopsis thaliana.
OS
XX US5773691-A.
PN
XX 30-JUN-1998.
PD
XX 07-JUN-1995; 95US-0474633.
PF
XX 19-MAR-1992; 92US-0855414.
PR 18-MAR-1993; 93WO-US02480.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.
PA
XX Falco SC, Keeler SJ, Rice JA;
PI WPI; 1998-387117/33.
DR N-PSDB; AAV35857.
XX
XX Chimeric genes encoding lysine production enzymes - useful for
PT increasing transgenic seed lysine content without being inhibited by
PT high levels of the amino acid
XX
XX Example 20; Columns 155-156; 106pp; English.
PS
XX The present sequence represents a protein homologous to fungal
CC saccharopine dehydrogenases. The specification describes a
CC Corynebacterium dapsa gene, which encodes a dihydrodipicolinic acid
CC synthase (DHDPs) enzyme, which was used to create chimeric genes of
CC the invention. The chimeric genes contain a nucleic acid fragment
CC encoding a DHDPs enzyme which is insensitive to inhibition by lysine
CC operably linked to a plant chloroplast transit sequence and to a plant
CC seed-specific regulatory sequence. The chimeric genes are useful for
CC producing plants containing increased levels of lysine, especially in
CC corn (Zea mays) and soybean (Glycine max).
XX
SQ Sequence 74 AA;

Query Match 5.1%; Score 270; DB 19; Length 74;
Best Local Similarity 67.6%; Pred. No. 2.5e-16;
Matches 50; Conservative 18; Mismatches 6; Indels 0; Gaps 0;

Qy 948 KHOATLLEFGKVENGRSTTAMALTVGIPAAIGALLLNKVKVOTKGVIRPLQPEIYVPALE 1007
Db 1 khtatllefkdikngqttamaktvgipaalgalliedkiktrgvlrpleaeayvpald 60

Qy 1008 ILESSGIKLVEKVE 1021
Db 61 ilqaygiklmeakae 74

RESULT 14
AAW87758
ID AAW87758 standard; Protein: 74 AA.
XX
XX AAW87758;
AC

XX 29-MAR-1999 (first entry)
DT
XX Arabidopsis lysine ketoglutarate reductase.
XX
DE
XX Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW transgenic plant; seed.
KW
XX Arabidopsis thaliana.
OS
XX WO9842831-A2.
PN
XX 01-OCT-1998.
PD
XX 27-MAR-1998; 98WO-US06051.
PF
XX 27-MAR-1997; 97US-0824627.
PR
XX (DUPO) DU PONT DE NEMOURS & CO E I.
PA
XX Epelbaum SU, Falco SC, McDevitt RE;
PI WPI; 1999-045139/04.
DR N-PSDB; AAV99554.
XX
XX Nucleic acids and chimeric genes for increasing seed lysine content
PT - comprise sequence encoding all or part of lysine ketoglutarate
PT reductase, useful to improve nutritional quality of seeds from
PT transformed plants
XX
XX Claim 2; Page 166-167; 23lpp; English.
PS
XX This is the amino acid sequence of an Arabidopsis thaliana
CC lysine ketoglutarate reductase (LKR) polypeptide, identified on the
CC basis of homology to a fungal glutamate-forming saccharopine
CC dehydrogenase. Isolated nucleic acid fragments comprising
CC sequences encoding all or part of plant LKR polypeptides are new.
CC Also claimed are: (1) a chimeric gene comprising the fragment (or a
CC subfragment) operably linked to a suitable seed specific regulator,
CC where the chimeric gene reduces LKR activity in plant seeds
CC transformed with it; (2) plant cells and seeds in which LKR
CC activity is reduced due to a mutation in the gene encoding LKR or
CC transformation with the chimeric gene; (3) a nucleic acid fragment
CC comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC synthase (DHDPs), substantially insensitive to lysine inhibition is
CC operably linked to a plant chloroplast transit sequence and to a
CC plant seed-specific regulatory sequence; (4) plants comprising in
CC genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC from (3). The chimeric genes can be used to produce plant cells
CC and seeds with reduced LKR activity, especially in Arabidopsis,
CC corn, soybean, rapeseed, wheat and rice (claimed). LKR is
CC important in controlling free lysine accumulation in plant seeds.
CC LKR activity reduction may be achieved by cloning the claimed
CC fragment, preparing a chimeric gene for cosuppression of LKR,
CC expression of antisense RNA for LKR, and transforming plants with
CC the chimeric gene.
XX
SQ Sequence 74 AA;

Query Match 5.1%; Score 270; DB 20; Length 74;
Best Local Similarity 67.6%; Pred. No. 2.5e-16;
Matches 50; Conservative 18; Mismatches 6; Indels 0; Gaps 0;

Qy 948 KHOATLLEFGKVENGRSTTAMALTVGIPAAIGALLLNKVKVOTKGVIRPLQPEIYVPALE 1007
Db 1 khtatllefkdikngqttamaktvgipaalgalliedkiktrgvlrpleaeayvpald 60

Qy 1008 ILESSGIKLVEKVE 1021
Db 61 ilqaygiklmeakae 74

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 15:44:58 ; Search time 26.05 seconds
(without alignments)
3769.800 Million cell updates/sec

Title: US-09-049-304A-122

Perfect score: 5286

Sequence: 1 CARLLGGGKNGPRVNRIV.....VPALETLESSGIKLVKVFET 1022

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71.*

1: piri.*

2: piri2.*

3: piri3.*

4: piri4.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4750	89.9	1056	2 T02930	lysine-ketoglutarate
2	3411	64.5	1064	2 T05195	saccharopine dehyd
3	1429.5	27.0	934	2 T15063	hypothetical prote
4	716	13.5	450	2 T40337	probable saccharop
5	672.5	12.7	446	2 S41937	saccharopine dehyd
6	608	11.5	177	2 T07843	probable lysine-ke
7	216.5	4.1	352	2 A71176	hypothetical prote
8	172	3.3	369	2 A36467	saccharopine dehyd
9	157	3.0	368	2 T50174	probable saccharop
10	154.5	2.9	373	2 S48496	saccharopine dehyd
11	148.5	2.8	366	2 F97474	hypothetical prote
12	148.5	2.8	366	2 AB2893	dehydrogenase Atu0
13	145.5	2.7	1825	2 S13507	microtubule-associ
14	141.5	2.6	605	2 A37981	microtubule-associ
15	139.5	2.6	605	2 I39837	dnak-type molecula
16	136.5	2.6	414	2 A82177	conserved hypothet
17	136.5	2.6	703	2 AC2430	hypothetical prote
18	135	2.5	405	2 B75420	conserved hypothet
19	133.5	2.5	1300	2 T00317	conserved serine pr
20	132	2.5	398	2 S74347	hypothetical prote
21	130.5	2.5	419	2 D97974	conserved hypothet
22	130	2.5	407	2 E69409	conserved hypothet
23	127.5	2.4	411	2 G69215	conserved hypothet
24	127.5	2.4	419	2 B95106	conserved hypothet
25	127	2.4	705	2 S76729	hypothetical prote
26	127	2.4	1068	2 T48756	mitochondrial nico
27	125.5	2.4	1451	2 A36468	SPT6 protein - yea
28	125	2.4	2261	2 T20978	hypothetical prote
29	124	2.3	5255	2 T31677	bacitracin synthet

30	123	2.3	408	2 C69448	conserved hypothet
31	122	2.3	1311	2 T08986	hypothetical prote
32	122	2.3	4717	2 T41581	hypothetical colle
33	121	2.3	760	2 T34414	hypothetical prote
34	121	2.3	2672	2 A48126	translation activa
35	120	2.3	493	2 A26941	cholesteryl ester
36	120	2.3	2241	2 T20971	hypothetical prote
37	119	2.3	401	2 C81435	hypothetical prote
38	119	2.3	1828	2 A40115	microtubule-associ
39	118.5	2.2	360	2 S74638	alanine dehydrogen
40	118.5	2.2	384	2 AG1948	hypothetical prote
41	118	2.2	428	2 A45732	phosphopruvate hy
42	117.5	2.2	946	2 A71805	probable ATP-depen
43	116.5	2.2	399	2 H71810	hypothetical prote
44	116.5	2.2	1450	2 S78060	probable DNA-direc
45	116.5	2.2	2802	2 F97686	cyclic beta-(1-2)

ALIGNMENTS

RESULT 1

T02930

lysine-ketoglutarate reductase / saccharopine dehydrogenase (NADP+, L-glutamate-formi
N:Contains: lysine-ketoglutarate reductase; saccharopine dehydrogenase (NADP+, L-glut
C:Species: Zea mays (maize)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T02930

R:Cord-Neto, G.; Kemper, E.L.; Arruda, P.

submitted to the EMBL Data Library, May 1998

A:Description: Lysine-ketoglutarate reductase/saccharopine dehydrogenase bifunctional

A:Reference number: Z14777

A:Accession: T02930

A:Status: translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-1056 <COR>

A:Cross-references: EMBL:AF003551; NID:Q3157909; PID:G3157910

A:Experimental source: cultivar AGROCERES F-352 commercial hybrid

C:Keywords: oxidoreductase

Query Match 89.9%; Score 4750; DB 2; Length 1056;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 941; Conservative 18; Mismatches 58; Indels 6; Gaps 5;

QY	1	CARLLGGGKNGPRVNR	IIVQPSTRRIHDAQYEDAGEISEDLSECLLIIGIKQPKLQM	60
Db	39	CARLLGGGKNGPRVNR	IIVQPSTRRIHDAQYEDAGEISEDLSECLLIIGIKQPKLQM	98
QY	61	ILSDRAVAFPSHTHKAQ	KNPLLDKTLERVSFLDYELIVGGDGKSLAFGKFAAGL	120
Db	99	ILSDRAVAFPSHTHKAQ	KNPLLDKTLERVSFLDYELIVGGDGKSLAFGKFAAGL	158
QY	121	IDFLHGLGQRYLSLGS	YSTPFLSLGQSHMYPSLAAKAAVIVAAEETATFGLPSGICPIVF	180
Db	159	IDFLHGLGQRYLSLGS	YSTPFLSLGQSHMYPSLAAKAAVIVAAEETATFGLPSGICPIVF	218
QY	181	VFTGVNVSOGAEIEFK	LPLPHTFVDAEKLPEIFQARNLSKQSQSTRKRVFOLYGCWVTSRD	240
Db	219	VFTGVNVSOGAEIEFK	LPLPHTFVDAEKLPEIFQARNLSKQSQSTRKRVFOLYGCWVTSRD	278
QY	241	IVSHKDPTRQFDKDY	YAHPEHYTPVFHERIAPYASIVNCMYWEKFRFPPLLNMDQLQL	300
Db	279	MVSHKG-SHOTLTSD	YIAHPD-TTPCFHERIAPYASIVNCMYWEKFRFPPLLNMDQLQL	336
QY	301	METGCPVLGVCDITC	DIGGSIEFINKSTSIERFFRYDPSKNSYHDDMEGAVVCLAVDI	360
Db	337	METGCPVLGVCDITC	DIGGSIEFINKSTSIERFFRYDPSKNSYHDDMEGAVVCLAVDI	396
QY	361	LPTEFSKEASQHPGN	ILSRVLASVSKQPAELPSYLRRACIAHAGRLTPLYEYIPMRN	420
Db	397	LPTEFSKEASQHPGN	ILSRVLASVSKQPAELPSYLRRACIAHAGRLTPLYEYIPMRN	456

hypothetical protein R02D3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #tex
R:Mu, X.; Antoniou, B.
C:Accession: T15063
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid R02D3.
A:Reference number: Z18284
A:Accession: T15063
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-934 <WUX>
A:Cross-references: EMBL:AF038615; NID:g2736322; PID:g2
A:Experimental source: strain Bristol N2; clone R02D3
C:Genetics:
A:Gene: CESP-R02D3.1
A:Map position: 4
A:Introns: 35/2; 135/3; 262/1; 304/3; 340/3; 476/2; 555/

Query Match	27.0%	Score	1429.5	DB 2	Length	934			
Best Local Similarity	32.5%	Pred. No.	4.6e-88						
Matches	334	Conservative	190	Mismatches	321	Indels	183	Gaps	24
QY	10	KNGVRNRIIVQPSRRIRHHDAQVEDACCEISEDLSECGLLIGIKQPKQWLMSDRAYAF	69						
DB	58	KSG--VN-VLQPSNRRAYPIQDISAGAIYREDLSEAHIIIMSVKSVPIQDLINKTVAF	114						
QY	70	FSHTHKAQKENMPDLDKTLEERSVLFDFYELIVGDGKRSLAFAGFAGAGLIDFLHGLGQ	129						
DB	115	FSHTIKAQQDNEMDLTILQNRILLDYEKCDKGRKRVMGKAWGAGFIDILHIGL	174						
QY	130	RYLSIGSTPPLSGQSHMYPSSLAAAKAAVIVAAEIIATFGLPSGICPIVFVFTGVGNVS	189						
DB	175	RLLAGHTPFLHMLAHNYNDSHAINALRDIGYEIALDRMPRSLGLPIFVFTGSGNV	234						
QY	190	QGAQIEFKLLPHTFVDAEKLPEIFQARNLSKQSQSTRVQLYGCVVTSRD--IVSHKDP	247						
DB	235	QGARELEHPLHEVVDVATLPKVAQKQOLNK-----VYGCVVTRKDHFPVKHGGP	284						
QY	248	TRQDKGDYVAHPHYTVFHERIAPYASVIVNCYWEKFRPPLLNMDQLQML-----	301						
DB	285	---FDKKEFEQFPDRYTSKATEIAPVASVIINGVYNDQASPRLIITIPDAKNLLTPVQRY	341						
QY	302	-ETGCP-----LVGVCDDITCDIGSIEFINKSTSIERPFRYPDSKNSYHDDMEGA--GVV	354						
DB	342	DTGCPPTLPHLITALCDISADPGSVSEFRECITDKPFAIYDADFTSDSDFDAASGCL	401						
QY	355	CLAVIDILPTFSKASQHGNIILSRVLASLVASKQP-----AELPSVLRRACTAHAGRLT	409						
DB	402	VCSIDNMPAQMPIEATFQGNLLYPWLWLLNTSNDQHFDRLCQRTKTEIKNAITTDQCKLT	461						
QY	410	PLYEYTPRMNTMIDLAPAKTNPLDPKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFH	469						
DB	462	PNFEYTAQLRKDKKAASANSRVN-----	485						
QY	470	LVRCEVGQSTDDNSYSELEVGADDTATLDKIIDLSTSLANEHGGDHDQAGEIELALIKG	529						
DB	486	-----GG-----	487						
QY	530	VNEYETDVTIDKGPKILILGAGRCVRPAAEFLASYPDICTYGVDDHDADQIHIVIASLY	589						
DB	488	-----TTDK---RVLLLAGWGSPPADFYSKKQ-----VNLTVATES	523						
QY	590	QKDAEETVDGIENTATQLDVA--DIGSLDLSVQVEVWISLLPASPFAHAATAGVCIELKKH	648						
DB	524	QRDQGRLCITS--PNIQSVVVDIARESHTMERLIREHLDVLSLLPFNFHPLVAKKICISNQRD	582						
QY	649	MVTASVDESMSLSQAQADAGVYITLCEMGLDPGIDHLMKMKWIDAHARKGKIKAFYSY	708						
DB	583	MVTSSVYSPELEALDRAANDADVTINNEAGLDPGIDHMLAMECFDDIKEHGGRITYSF	642						
QY	709	CGGLPSPAANNPLAYKFSWNPAGALRSKGNPAVYKELGETIHVDGNHLYESAKRLRLRE	768						

Db	643	CGGLPAPWSONPLRYKFSWSPKGVLTALMNPALKNGKIVEPVGSVVDNL--IDIDF	700
QY	769	LPFALEHLPNRNSLIYGDLYGSKAEASTIYRATRYEGFSEINVTLSKTGFDDAANHPL	828
Db	701	MFGLNLGIFFPNRDS TKYSDVYGLGNDCKTIIIRGTLRYOGFVDTKALHSVGLLSGDNIDS	760
QY	829	LQDTSRP--TVKGFLELDNNISINTIDLDIEASGGYDDDLIARLLKLGCCCKNKEITAVKT	886
Db	761	FTSNGIPDLTWK----ELIASLS--NQKLDI-----FPDSLRIHI-----EEKVGGKG	802
QY	887	VKTIKFGLHHEETQIPKGCSSPFVICQRMQRNAYGHNEQDMVLLHHEVEVEYDPGQA	946
Db	803	LSAENLGLFSDKVVDRH-GTPIDTLAQYLAKILAFKDHESDLVNLNHDIGAQLPGCN-S	860
QY	947	EKHQATLEFCKVNGRSTTAMALTVCIGPAAGIALLKLNKVTGKVIIRPLOPEIIVPAL	1000
Db	861	ERHRISLVQYG--NPNGFSAMARTVGYTTAIVSHVMLNNEIQRAGIQRPILKEVYRPAL	917
QY	1007	EILESGI 1014	
Db	918	KRLRDFGI 925	
RESULT	4		
T40337			
		probable saccharopine dehydrogenase (NADP+, L-glutamate-forming) (EC 1.5.1.1)	
		C:Species: Schizosaccharomyces pombe	
		C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Feb-2000	
		C:Accession: T40337	
		R:Lyne, M.; Rajandream, M.A.; Barrel, B.G.; Beck, A.; Reinhardt, R.; Pohl,	
		submitted to the EMBL Data Library, March 1998	
		A:Reference number: Z21921	
		A:Accession: T40337	
		A:Status: preliminary; translated from GB/EMBL/DBJ	
		A:Molecule type: DNA	
		A:Molecule type: DNA	
		A:Residues: 1-450 <LYN>	
		A:Cross-references: EMBL:AL022244; PIDN:CAAL8292.1; GSPDB:GN00067; SPDB:SPDB	
		A:Experimental source: strain 972h-; cosmid c3B8	
		C:Genetics:	
		A:Gene: SPDB:SPB3B8.03	
		A:Map position: 2	
		C:Keywords: oxidoreductase	

Query Match	13.5%;	Score 716;	DB 2;	Length 450;
Best Local Similarity	35.1%;	Pred. No. 2e-40;		
Matches 168;	Conservative 93;	Mismatches 182;	Indels 36;	Gaps 11;
Qy	544	PKILTLGAGRCVCPAAEFLASYPDICTYGVDDHDADQIHHVIVASLYQKDAETVDGIENT	603	
Db	2	PSILLGSGFVAHPPTLEYLSR-----KKNNTIVACHTLSKAEAFINGPNS	48	
Qy	604	TATQDLDADIGSLDVSQVEVVISLLPASFHAAIAGVCIELKHKMVTASYVDESMNLS	663	
Db	49	KAIALDVNDEAALEKAVSEHDTLSLIPTVTHATVYKAAIKHGKIVCTTSYVNPMAELE	108	
Qy	664	QAANDAGVTIICEMGLDPGIDHLSMKMWIDBAHARKGKIKAFSTCYGGLSPAAANNPLA	723	
Db	109	EAAIKAGSICHNEIGVDPGIDHLYAIKTIEEVHAGKGIKSFSLCYGGLPAPEDSNPNLG	168	
Qy	724	YKFSWNPAGALRSGKNPAAVYFLGTHIVDGHNLVYESAKRLRLRELPAFALEHLPNRNSL	783	
Db	169	YKFSWSRGVLLARNSSAKFYENGLBGDKMLMETAKPYFI--YPGATFVCYPNRDS	226	
Qy	784	IYDGLYGISKEASTYRATXRYEGFSEIMWTLSTKTFDF-AANHPLLDQTSRPYKGFELD	842	
Db	227	VYQERYQI-PEAETIIRGTLRYQGFPPEFHCLVDWGMFLDETAQEVLSPEAPALPWK----	281	
Qy	843	ELLNNISTINTDLOIEASGGVDDDLIARLLKLGCKCKNKEIAKVTKTKFGLGHEETQI-	901	
Db	282	EVTARV-----IKAESSEADLIIKSIIRHFKDDDDKKRLLINGLKLMGFESSKPYT	333	

QY	902	PKCCSPFDVICORMEORMAYGHNQDMVLLHHEVEVEYPDGQPAEKHQATLLRFGKVEN	961
Db	334	PRG--NPLDTLCATLELMQYEEGERDMLILQHKFEVETKEG-RQTRCTLLDITG-VPN	389
QY	962	GRSTTAMALTGIPAAIGALLLKNKQVTKGVRILQPIYVPALFLEISGKILVEKY	1020
Db	390	G--YTSNAKLVGPGVATQIILDGVINTPGVLANDMKLCGLPLDTLAKEGIRLEEEI	446
RESULT	5		
S41937		saccharopine dehydrogenase (NADP+, L-glutamate-forming) (EC 1.5.1.10) - yeast (Saccharom	
N:Alternate names:		protein N3461; protein YNR050C	
C:Species:		Saccharomyces cerevisiae	
C>Date:		03-May-1994 #sequence_revision 27-Jan-1995 #text_change 20-Jun-2000	
C:Accession:		S41937; S63381	
R:Feller, A.		submitted to the EMBL Data Library, January 1994	
A:Reference number:		S41936	
A:Accession:		S41937	
A:Molecule type:		DNA	
A:Residues:		1-446 <PEL>	
A:Cross-references:		EMBL:X77363; NID:g453185; PID:g453186	
R:Pohl, T.M.		submitted to the Protein Sequence Database, April 1996	
A:Reference number:		S63346	
A:Accession:		S63381	
A:Molecule type:		DNA	
A:Residues:		1-446 <POH>	
A:Cross-references:		EMBL:X71665; NID:g1302563; PID:g1302564; MIPS:YNR050C	
A:Experimental source:		strain S288C	
C:Genetics:			
A:Gene:		SGD:LYS9; LYS13	
A:Cross-references:		SGD:S0005333; MIPS:YNR050C	
A:Map position:		14R	
C:Keywords:		NADP; oxidoreductase	
Query Match		12.7%; Score 672.5; DB 2; Length 446;	
Best Local Similarity		35.1%; Pred. No. 1.7e-37;	
Matches		170; Conservative 82; Mismatches 183; Indels 49; Gaps 13;	
QY	543	GPKTILGAGRCVCPAAEFILASYPDICTYGVDDHDADQIHVIVASLYQKDAETVDGIE	602
Db	2	GKNVLLGSGFVAGPVIDTLAA-----NDDINVTACRTLANAQAALAKP-SG	47
QY	603	TTATQDQVADIGSLSDLSQVEVVISLLPASHAAIAGVCIELKHKMVTASYVDESMS	662
Db	48	SKAISLDVTDSDALDKVLADNDVVISLPTTFHPNVVKSARKTKDVTSSYISPALREL	107
QY	663	SQAARKAGVTILCEWGLDPGIDHLSMKMIDEAHARKGIKAFTSYCGGLPSPAAANNPL	722
Db	108	EPEIVKAGITVMNEIGLDPGIDHLYAVKTTIDEVHRAGGKLKSLFSYCGGLPAPEDSDNPL	167
QY	723	AYKFSWNPAGALRGSKNPANYKFFETIHYVDGHNLYESAKRLRLRELPAFALEHLPNR	782
Db	168	GKFSWSRGRVLLARNSAKYWKDKGKETVSSDLMATAKPYFT--YPGAFVCYPNRDS	225
QY	783	LIYGLDYGISKEASTIVRATXRYEGFSEIMVTLSKTGFF-DAANH-----PLQDTSRPTY	837
Db	226	TIFKDLHYI-PEAETVIRGTILRYOGFFPVFKALVDMGLKDDANEIFSKPIANNEALKQY	284
QY	838	KGFLDELNNISTINTDLDIASGGYDDDLIARLLKLGCCCKNKEIAVTKTKIFLGHE	897
Db	285	LG-----AKSTSKEDLIASIDSKATWKDDREDREIRILSGFAWLGLFS	325
QY	898	ETQI-PKCCSPFDVICORMEORMAYGHNQDMVLLHHEVEVEYPDGQPAEKHQATLLEF	956
Db	326	DAKITPRG--NALDPLCARLELMQYEDNERKDMVQLQHKFGIEWADG-TTETRTSLVDY	382
QY	957	GKVENGRSTTAMALTGIPAAIGALLLKNKQVTKGVRILQPIYVPAL-EILESSGK	1015
Db	383	GKV-CGYS--SMAATVGYGVAIATKFKVLDTGKIPGLIAPYSPEINDPIMKELDKYGIY	439

QY	1016	LVEK 1019	
Db	440	LKEK 443	
RESULT	6		
T07843		probable lysine-ketoglutarate reductase / saccharopine dehydrogenase (EC 1.5.1.-) - r	
N:Contains:		lysine-ketoglutarate reductase; saccharopine dehydrogenase	
C:Species:		Brassica napus (rape)	
C>Date:		14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999	
C:Accession:		T07843	
R:Deleu, C.; Coustaut, M.; Niogret, M.F.; Larher, F.		submitted to the EMBL Data Library, January 1998	
A:Reference number:		216169	
A:Accession:		T07843	
A>Status:		preliminary; translated from GB/EMBL/DBD	
A:Molecule type:		mRNA	
A:Residues:		1-177 	
A:Cross-references:		EMBL:AF042184; NID:g2809206; PIDN:AAB97685.1; PID:g2809207	
A:Experimental source:		cv oleifera	
C:Keywords:		oxidoreductase	
Query Match		11.5%; Score 608; DB 2; Length 177;	
Best Local Similarity		68.6%; Pred. No. 8.5e-34;	
Matches		118; Conservative 24; Mismatches 26; Indels 4; Gaps 2;	
QY	570	TY-GVDDHDADQIHVIVASLYQKDAETVDGIENTATQDQVADIGSLSDLSQVEVVIS	628
Db	9	TYXGGEQRD---VRVIVASLYLKDAKEVGEVPEAVQLDVSDESLLKYVSEVDVVL	65
QY	629	LLPASFAHAAIAGVCIELKHKMVTASYVDESMSNLSSQAADAGVTILCEMGLDPGIDHLS	688
Db	66	LLPASCHASVAKTCTIELKKHLITASVDDETSLGHEKAKHAGITILGEMGLDPGIDHMA	125
QY	689	MMIDEAHARKGIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRGSKNP	740
Db	126	MKMINAEHARKGVKFSYCGGLPSPAAANNPLAYKFSWNPAGAIRAGNP	177
RESULT	7		
A71176		hypothetical protein PH1888 - Pyrococcus horikoshii	
C:Species:		Pyrococcus horikoshii	
C>Date:		14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000	
C:Accession:		A71176	
R:Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Hino, Y.; Yamamoto, S.; Se		M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu	
DNA Res. 5, 55-76, 1998			
A:Title:		Complete sequence and gene organization of the genome of a hyper-thermophil	
A:Reference number:		A71000; MUID:98344137	
A:Accession:		A71176	
A>Status:		preliminary; nucleic acid sequence not shown; translation not shown	
A:Molecule type:		DNA	
A:Residues:		1-352 <KAW>	
A:Cross-references:		GB:AP000007; NID:g3236134; PIDN:BAA30800.1; PID:g3258117	
A:Experimental source:		strain OF3	
A:Note:		this accession replaces an interim accession for a sequence replaced by GenBa	
C:Genetics:			
A:Gene:		PH1688	
Query Match		4.1%; Score 216.5; DB 2; Length 352;	
Best Local Similarity		23.2%; Pred. No. 6.9e-07;	
Matches		89; Conservative 63; Mismatches 149; Indels 83; Gaps 14;	
QY	545	KILTLGAGRCVCPAAEFILASYPDICTYGVDDHDADQIHVIVASLYQKDAETVDGIENT	604
Db	5	KVLLIAGNIGRAIWDL-----KDEFDVYIGDVNNENLEKVK---EFAT	46
QY	605	ATQDQVADIGSLSDLSQVEVVISLLPASFAHAAIAGVCIELKHKMVTASYVDESMSNL	664

Db 47 PLKVDASNFDKLVEMKFEELVIGALPGFLGFKSIKAAIKSKVDMVDVSEMPENPLELRD 106
QY 665 AAKDAGVTILCENGLDPCIDHLSMKMKIDAHARKKIKRAFTSYCGGLSPSAAANPLAY 724
Db 107 EAEKAQVTIVFDAGFAPGSLNMLGRIFQELDLKEGYI-----YVGLPKD--PKPPLY 159
QY 725 KFSNPAGALRSKGNPAVKYKFLGETHVDGHNLYESAKRLRLR--ELPAPALEHLPNRNS 782
Db 160 KITWSPDLIEETRPARYIRNCKSVKVDP---LSEVKVKIKGKFEFAFISDGLRSMLE 216
QY 783 LIYGLYIGISKEASTYRATYRVEGSEIMVTLSTKTFDDAAN-----HPLLQ---- 830
Db 217 TI-----NSERLEEWTLRMPGHLEKIKVLRELGFEPENLDTLRAVIEPLMYENK 267
QY 831 --DTSRPTVKG-----FL---DELLANNISPTINDLIEASGGYDDDLIARLLKLGCC 877
Db 268 DFSIMVKVKGEGEMEFFLYDEDSMFSSMRVT-----GFTAAIISRIVAENTC 318
QY 878 KKEIAVTKVTI--KFLGLHEET 899
Db 319 -----TFGVIPPEILGMREDT 334

RESULT 8
A36467
saccharopine dehydrogenase (NAD+, L-lysine-forming) (EC 1.5.1.7) - yeast (Yarrowia lipolytica)
C:Species: Yarrowia lipolytica, Candida lipolytica
C:Date: 15-Feb-1991 #sequence_revision 15-Feb-1991 #text_change 21-Jul-2000
C:Accession: A36467
R:Xuan, J.W.; Fourrier, P.; Declerck, N.; Chasles, M.; Gaillardin, C.
Mol. Cell. Biol. 10, 4795-4806, 1990
A:Title: Overlapping reading frames at the LYS5 locus in the yeast Yarrowia lipolytica.
A:Reference number: A36467; MUID:90355996
A:Accession: A36467
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-369 <XU>
A:Cross-references: GB:M34929; NID:gl73260; PIDN:AAA35248.1; PID:gl73262
C:Keywords: NAD; oxidoreductase

Query Match 3.3%; Score 172; DB 2; Length 369;
Best Local Similarity 20.5%; Pred. NO. 0.00075;
Matches 81; Conservative 64; Mismatches 159; Indels 92; Gaps 12;
QY 18 IIQVPSRRIHDAQYEDAGCEISDLS-----ECGLIIGIKOPKQLMILSDRAYFTSH 72
Db 37 VFVEKSPRIFDQDFVDGATLVGEGSVSAPEDRMIIIGLKELPEESFPISHEHIQFAH 96
QY 73 THKAQENMPLDKLILEERVSFLDYELIVGDDGKRSLAFGRFAGRAGLIDFLHGLGQRYL 132
Db 97 CYKQGGWKDVLRSFPAGNGTLYDLEFLDDNGRRVAAGFGHAGFAG-----A 144
QY 133 SIGYSTPFLSLGOSH-----MYPSLAAKAAVIVAAEETATFGLPGSGICPIVVF 182
Db 145 AIGVET--WAFQQTPODSENLPVGSAYEN---ETELVDKIKKDLAAAVKEGSKLPTVLVI 199
QY 183 TGVGNVSGOAGIEIKLLPHTFTVDAEKLPEIFOARNLSKOSOSTKRVFOLYGCVVTSRDIV 242
Db 200 GALGRCGSGAIDLARKV-----GIPE-----ENIIRWDMNEYK----- 232
QY 243 SHKDPTRQFDKGYAHPEHTVPFHRIAPYASVIVNCMYWEKRFPPPLNMDLOQLME 302
Db 233 -----KGGPF-----QEIAD-ADIFINCIVLSQPIPPFIYDILLNKEIR 270
QY 303 TGCPLVGVCDITCDIGGSIEFINKSTIERPFFRYDPSKNSYHDDMEGAGVYCLAVDILP 362
Db 271 KLSIVDVDSATTHPNPVPVYTTATFDHPTVPVET-----AGPKLSVCSIDLHP 322
QY 363 TFSKEASQHFNGILSRILVASLASKVQPAELPSYL 398
Db 323 SLLPREASEAFS---EALLPSLLQLPQDTPAVWTR 355

RESULT 9
T50174
probable saccharopine dehydrogenase [Imported] - fission yeast (Schizosaccharomyces fission yeast)
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: T50174
R:Zimmermann, W.; Wambutt, R.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, November 1999
A:Reference number: Z25036
A:Accession: T50174
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-368 <ZIM>
A:Cross-references: EMBL:AL131156; PIDN:CAB61467.1; GSPDB:GN00066; SPDB:SPAC227.18
A:Experimental source: strain 972h(-); cosmid c227
C:Genetics:
A:Gene: SPDB:SPAC227.18
A:Map position: 1

Query Match 3.0%; Score 157; DB 2; Length 368;
Best Local Similarity 20.6%; Pred. NO. 0.0076;
Matches 83; Conservative 63; Mismatches 156; Indels 100; Gaps 14;
QY 17 RIIVQPSRRIHDAQYEDAGCEISDLS-----ECGLIIGIKOPKQLMILSDRAYAFF 70
Db 35 QITIERSSQRAFKDKFEERLGFPMVPEGSWRHAPKDAVIIGLKELPENDNSPLKHTHIQF 94
QY 71 SHTHKAQENMPLDKLILEERVSFLDYELIVGDDGKRSLAFGRFAGRAG-----LIDFLHG 126
Db 95 AHYCNQNGWREVSFRFPAGNGLYDLEFLDDNGRRVAAGFYHAGFAGSIAISLWHAHQ 154
QY 127 LGORYLSLGYSTPFLSLGOSHMYPSLAA---AKAAVIVAAEETATFGLPGSGICPIVVF 182
Db 155 L-----LHPNKQFPAIRFPPEKSLVRHVRAQVRLAKKNNQIPRILVI 199
QY 183 TGVGNVSGOAGIEIKLLPHTFTVDAEKLPEIFOARNLSKOSOSTKRVFOLYGCVVTSRDIV 242
Db 200 GALGRCGGTGACDL-----ASKIGIPED-NILRWINDINETKK----- 233
QY 243 SHKDPTRQFDKGYAHPEHTVPFHRIAPYASVIVNCMYWEKRFPPPLNMDLOQLME 302
Db 234 --GGPFTEITSD-----IFVNCIYLSMPIPKFCVESLN---V 267
QY 303 TGCPLVGVCDITCDIGGSIEFINKSTIERPFFRYDPSKNSYHDDMEGAGVY-----CL 356
Db 268 PNRLRVVCDVSCD-----TTNPNNPIPIYVNTTDFHPTVEKGVTTTPPPLEVI 317
QY 357 AVDILPTEFSKEASQHFNGILSRILVASLASKVQPAELPSYL 398
Db 318 SIDHLPTLLPRESSEAFS---EALIPSLALLKQVDNAPVWVR 356

RESULT 10
S48496
saccharopine dehydrogenase (NAD+, L-lysine-forming) (EC 1.5.1.7) - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YIR034c
C:Species: Saccharomyces cerevisiae
C:Date: 31-Mar-1992 #sequence_revision 27-Jan-1995 #text_change 06-Feb-1998
C:Accession: S48496; S41936
R:Rowley, K.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48478
A:Accession: S48496
A:Molecule type: DNA
A:Residues: 1-373 <ROW>
A:Cross-references: GB:247047; EMBL:Z38061; NID:g603997; PID:g763379; MIPS:YIR034c
R:Teller, A.
submitted to the EMBL Data Library, January 1994
A:Reference number: S41936
A:Accession: S41936

```
Db 4 IVVIGAGNIGSAIAWMLAATGDY-RITVADRSDAQ-----LANVPAHERVD----- 4
```

Db 200 --SLDGVTEAFNTSSGLGTLCATLEGKVRTMNYRTIRYPGHVAIMKALLNDLNRNRD 25

A;Residues: 1-366 <KUR>
A:Cross-references: GR:AF007869: PTDN:AAK86751.1: PID:q15155947: GSPDB:GN00169

Query Match	2.8%;	Score 148.5;	DB 2;	Length 366;
Best Local Similarity	25.2%;	Pred. No. 0.028;		
Matches 105;	Conservative 42;	Mismatches 146;	Indels 123;	Gaps 21

Qy	546	ILILGAGRCVCPAAEEFLASYPDICTYGVDDHDADQIHVIVASLYOKDAEETVDGIENTTA	605
		: : : : : :	
		- - - - -	
Db	4	IWVGAGNIGSAIAWMUAATGY-RITVADRSADQ-----LANPVAHERVD-----	48

```

Qy 818  --TGFFDAANHPLQD-----TSRPTYKG-FLDELLNN-----IST-- 850
      | : || | | | | | | | | | | | | | | | | | | | | |
Db 258  VLKDLFENALPGTMQDWIVFVTCGTRNGRFLQETIYANKYVAGPYSGRMMSAQITTA 317
      | : || | | | | | | | | | | | | | | | | | | | | |
Qy 851  -INTDLDIASGGYDDDLTARLLKLGCCCKNKKEIAVTKTKFL-----GLHE 897
      | | | | | | | | | | | | | | | | | | | | | | | |
Db 318  GICTVDLLLAEGS-----LPQKGFVQEEVALP-----KFLERNFRGYGAHE 360
      | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
S13507
microtubule-associated protein MAP2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999
C:Accession: S13507
R:Marchal, D.; Delapierre, D.; Dresse, A.
A:Title: Cloning and partial sequencing of a new rat brain specific cDNA.
A:Reference number: S13507; MUID:89334524
A:Accession: S13507
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-1825 <MAR>
A:Cross-references: EMBL:X54100; NID:G56624; PIDN:CAA38034.1; PID:G56625
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1990
C:Superfamily: microtubule-associated protein.MAP2b; MAP2/tau repeat homology
C:Keywords: microtubule binding; tandem repeat
F:1667-1697/Domain: MAP2/tau repeat homology <MT1>
F:1698-1728/Domain: MAP2/tau repeat homology <MT2>
F:1729-1760/Domain: MAP2/tau repeat homology <MT3>

```

QY 739 NPAVYKELGTIIH---VDGHNLYESA---RLRLREL-PAPALEHLNPNRSLIYGDLY-- 789

Db 1121 -----DLVHQAEVDKESYSEGEHSLTMSLKPDEGKKTSPETSLIQDEVALK 1171

QY 790 -----GISKEASTI-VRATXRYGFSFIMVTLSTGTFDDAANPILL-QDTSRRTYK 838

Db 1172 LSVIEPCPPVPSADSSIDEKAEQMB-----FIQLPRESTETPDIPAIPSDVTQPOPE 1226

QY 839 GFIDELNNISTINTOLDIASGYD-----DDL-IARLLKLG-----CCKNKEIA 883

Db 1227 AVYSE---PAEVRGESEEIEAEGYDKLLPRSDTLQITDLLVPGSRFEFVETCPGEHGV 1283

QY 884 VKTVKTIK-----FLGHEETQIIPKGCSSPDVVICQRMEOQMAIGHNEDQM 929

Db 1284 VESVVTIEDFTIVVOTTTDDEGLGSH-----SVRFAAPVQPEERRRPHODELE 1334

QY 930 VLLHHEVEVEYPDGP-----AEKHQATLLEF 956

Db 1335 VLMAAEQAEPKDGSPDAPATPKBEVPFSEY 1366

RESULT 14

A37981

microtubule-associated protein 2b - rat

N:Alternate names: MAP2b

N:Contains: microtubule-associated protein 2c (MAP2c)

C:Species: Rattus norvegicus (Norway rat)

C:Date: 28-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 13-Aug-1999

C:Accession: A37981; B37981; SI0003; S07887; SI4568

R:Kindler, S.; Schulz, B.; Goedert, M.; Garner, C.C.

J. Biol. Chem. 265, 19679-19684, 1990

A:Title: Molecular structure of microtubule-associated protein 2b and 2c from rat

A:Reference number: A37981; MUID:91060576

A:Accession: A37981

A:Molecule type: mRNA

A:Residues: 1-1830 <KIN>

A:CROSS-references: GB:X51842; NID:956620; PIDN:CAA36135.1; PID:956621

A:Accession: B37981

A:Molecule type: mRNA

A:Residues: 1-151,1515-1830 <KI3>

R:Kindler, S.; Schwanke, B.; Schulz, B.; Garner, C.C.

Nucleic Acids Res. 18, 2822, 1990

A:Title: Complete cDNA sequence encoding rat high and low molecular weight MAP2.

A:Reference number: SI0003; MUID:90251471

A:Accession: SI0003

A>Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-1830 <KI2>

A:CROSS-references: EMBL:X51842; NID:956620; PIDN:CAA36135.1; PID:956621

R:Doll, T.; Papadrikopoulou, A.; Matus, A.

Nucleic Acids Res. 18, 361, 1990

A:Title: Nucleotide and amino acid sequences of embryonic rat MAP2c.

A:Reference number: S07887; MUID:90221819

A:Accession: S07887

A:Molecule type: mRNA

A:Residues: 1-151,1515-1830 <DOL>

A:CROSS-references: EMBL:X17682; NID:956622; PIDN:CAA35667.1; PID:956623

R:Matus, A.; Doll, T.

Submitted to the EMBL Data Library, May 1990

A:Reference number: SI4568

A:Accession: SI4568

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-476, 'H', 478-486, 'E', 488-525, 'R', 527-565, 'V', 667-670, 'K', 672-872, 'R', 874-875

A:CROSS-references: EMBL:X53455; NID:957619; PIDN:CAA37535.1; PID:957620

C:Genetics:

A:Gene: map2

C:Superfamily: microtubule-associated protein MAP2b; MAP2/tau repeat homology

C:Keywords: alternative splicing; microtubule binding; tandem repeat

F:1-1830/Product: microtubule-associated protein 2b #status predicted <MB2>

F:1-151, 1515-1830/Product: microtubule-associated protein 2c #status predicted <MC2>

F:1672-1702/Domain: MAP2/tau repeat homology <MT1>

F:1703-1733/Domain: MAP2/tau repeat homology <MT2>

F:1734-1765/Domain: MAP2/tau repeat homology <MT>
Query Match 2.7%; Score 141.5; DB 2; Length 1830;
Best Local Similarity 20.9%; Pred. No. 1.2;
Matches 167; Conservative 99; Mismatches 267; Indels 267; Gaps 41;
A:Title: Nucleotide sequence of a Bacillus megaterium gene homologous to the dnaK gen
A:Reference number: I39837; MUID:87231083
A:Accession: I39837
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-605 <RES>
A:Cross-references: GB:M31338; NID:g939628; PIDN:CAA68348.1; PID:g939629
C:Function:
A:Description: involved in protein folding and assembling/disassembling of protein co
A:Superfamily: heat shock protein 70
C:Keywords: ATP; molecular chaperone

Query Match 2.6%; Score 139.5; DB 2; Length 605;
Best Local Similarity 21.9%; Pred. No. 0.26;
Matches 143; Conservative 81; Mismatches 238; Indels 191; Gaps 30;
QY 428 AKTNPLPKKYSTLVSLSGHLFDKFLNEALDIETAGSGFHLVRCVQST-DMMS--- 483
Db 57 AITNP-----NTIISVRRHM-----GTDHKVEAEKGQYTPQEMSAII 93
QY 484 -----YSELEVGADDTATLQKIIDSLTSLAN--EHGGDHDAGQ-----EIE----- 522
Db 94 LOHLKGAEYLGEPVT-----KAVITVPAYFNDAERQATKDAGTAGLEVERIINEPTAA 149
QY 523 -LALKTGKVNVEYETDVTIDKGG-----PKLILGAGRCVCPAAEFSLASYPDICTYGVDDHD 577
Db 150 ALAYGLEKTDEDTVLVYDLGGGTFDVSILELGG-----VFVRATAGD-NRLGGDDFD 203
QY 578 ADQIHIVASLVQKD-----AETVDGIENTTATQLDVADIGSLSLDV 620
Db 204 QVILDYLAFAFKENGVDLSKDKMALORKDAEAKKDLGCVTSTQISLPI--TAGEA 261
QY 621 SQVEVVISLPSFAHAAGVCIELKHHMTASVYDESMNSLSQAQKAGVILCEMGLD 680
Db 262 GPLHEVLSRAKFDLSAGL-----VERTMAPVROALKDAG----- 298
QY 681 PGIDHLSMKMIDEAHARKGKIKAFKTSYCGGLPSPAAANNPLAYKFSWNPAGALRSKNP 740
Db 299 -----LSASELDK-----VLVGGSTRIPAVQDAIKKETGQDP-----HKGVPN 337
QY 741 AVYKFLGETIHDGHNLYESAKRLRLRELPAFAL-----EHLPNRNSLIYDLYCI 791
Db 338 DEVVALGAAL--QGGVLTGDKVDVLLDVTPLSLGTETMGVFTKLIERTTPTISKSQV 395
QY 792 SKEASTIYRATRY--EGFSEIMVTLSTKGFDAANHPQLQDTSRPTYKGFDELINNIS 849
Db 396 FSTAADSQTAVDTHVLQGERPMSADNKTGLRQLTDIP-----PAPRG-----VP 440
QY 850 TINTDLIDIASGGYDDDLIARLLKGCCKNKEIAVTKVTKIPLGLHHEETQIPKGCSSPF 909
Db 441 QIEVSFDIDKNGIVN-----VRAKDLG--TNKEQAI-TIKSSTGLSDDEIDRMVKEABENA 493
QY 910 DVTICQRMQRMAYGHNEQDMVLLHHEVEYDPQAPAEKHOATLLEF-GKVENGRTTAM 968
Db 494 DADQORKE-----EVELKNEADQLVFTTEKTLKDLGKVEAEVTKAN 536
QY 969 ALTVGTPAATGALLLNKVKQTKGRVPRLOPEIYVPALEILESSGKLVKEV 1021
Db 537 EAKDALKAATE-----KNDLEE---IKAKKDELQ-----EIVQALTVKLYEQAQ 577

Search completed: May 31, 2002, 15:47:40
Job time: 162 sec

Query Match 2.7%; Score 141.5; DB 2; Length 1830;
Best Local Similarity 20.9%; Pred. No. 1.2;
Matches 167; Conservative 99; Mismatches 267; Indels 267; Gaps 41;
QY 252 DKGDYAHPEHYTPVPHERIAPIYASVIVNCWYWEKRFPLLNMDQLQQLMETGCCPLVGC 311
Db 743 DEGDYLPPT--TPAVEK-----IPCPIESK-----EEDKTEQAKVTGGQTTQV- 786
QY 312 DITCDIGGSTEFINKSTIERPFRYDPSKNSYHDDMGAGVVCVLAVIDLPTFESEASQ 371
Db 787 -----ETSSSESP-----PAKEYKN-----GTV-MAPD-LPEMLDLAGR 820
QY 372 HFGNLSRLVASLAKVQPAELPSYLRRACIAHAGRLTPLYEYIPMRNMTIMDLAPAKTN 431
Db 821 -----SRL-ASVSADAENARRKSVSEAVVAESS-----TGLPPVADD 857
QY 432 PLPKKYSTLVSLSGHLFDKFLI---NEALDIETAG--GSFHLVRCVQSTD----- 480
Db 858 SOPVKPDSOLEDMGYCFNFKYTPPLSPVQDSNLSGSEGSFY-----EGDDKVRDD 910
QY 481 --DMSYSELEVG-----DDTATLQKIIDSLTS-----LANEHGGDHDAGQETIELAL 525
Db 911 LATDLSLIEVKLAAGRVKDEFTAEKASPPSSADKSGLSREFDQDKRANKDLDTVLEKS 970
QY 526 -----KIG-KYNEYETDVTIDKGGPKIILGAGRCVCPAAEFSLASYPDICT 570
Db 971 EEHVDSKEHAKSESEVGDKVLFGLVGYEQTSAKELITTKETAPERAERKGLSVPEVAE 1030
QY 571 YGVDDHDADQIHIVASLVQKDAETVDGIENTTATQLDVADIGSLSLDVSVQVEVVISL 630
Db 1031 VETTT-KADQGLDVA--KDDQSPID-----IKVSDFGOM----- 1063
QY 631 PASFAHAAGVCIELKHHMTASVYDESMNSLSQAQKAGVILCEMGLDPLGIDHLSMK 690
Db 1064 -ASGMSVDAGKTIELKFE-----VDQQLTSLSEAPQET-----DSFMG-- 1100
QY 691 MIDEAHARKGKIKAFKTSYCGGLPSPAAANNPLAYKFSWNPAGALRSKNPVYKFLGETI 750
Db 1101 -IESSHVDGAKVSETEVEKAVP-----DLV 1127
QY 751 H---VDGHNLYESA---KRLRLREL-PAFALHLPNRNSLIYDLY-----GIS 792
Db 1128 HQEAVDKEESYESSGEHESLWESLKDPEGKKETSPETSLIQDEVALKLSVEIPCPPVPS 1187
QY 793 KEASTI-YRATRYEGFSEIMVTLSTKGFDAANHPQL-QDTSRPTYKGFDELINNIS 850
Db 1188 EADSSIDEKAEVOME-----FIQLKEESTETPDIPAIPSDVTOPOPEAVVSE---PAEV 1239
QY 851 INTDLIDIASGGYD-----DDL-IARLLKLG-----CKNKEIAVTKVTKIK 891
Db 1240 RGEIEIEAEGEYKLLFRSDTLQITDILLVPGSREEFVETCPGEGHKGVVESVWTIEDFI 1299
QY 892 -----FGLHHEETQIPKGCSSPDTVCQRMQRMAYGHNEQDMVLLHHEVEVEYP 941
Db 1300 TVVQITTDGELGSH-----SVRFAAPVQPEERRPYPHDELEVLMAAQAEPK 1350
QY 942 DGQP-----AEKHOATLLEF 956
Db 1351 DGSPDAPATPEKEEVPFSEY 1370

RESULT 15
I39837
dnaK-type molecular chaperone - Bacillus megaterium
N:Alternate names: major heat shock protein dnaK homolog
C:Species: Bacillus megaterium
C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 21-Jul-2000
C:Accession: I39837
R:Sussman, M.D.; Setlow, P.L.
Nucleic Acids Res. 15, 3923, 1987
